Training the military engineer: a study of assessment and its validity

Thesis

Doctor of Education (EdD)

Training the Military Engineer: A Study of Assessment and Its Validity

Written by

Donald Andrew John Brooks

M1007616

&

Submitted on 10 December 2001

Contents

Chapter 5 - Presentation and Analysis of the Data

Chapter 6 - Conclusions

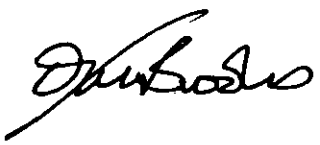
Chapter 7 - An Evaluation of My Study

Tables:

Appendices: Page 154

Statement on the Conduct of the Research

This dissertation has been produced from the individual research conducted throughout by myself, the undersigned. At no time has this dissertation, or any part herein, been submitted for another degree or other qualification of the Open University or any other university or institution.

Signed:  Dated: 10 December 01

Acknowledgements

My first acknowledgement and thank you must go to

Ginny

for her support over the years. Her unwavering support, encouragement, patience and understanding have been invaluable.

I would also like to thank the following:

Professor Robert McCormick for his advice, guidance and support as my tutor-supervisor throughout the doctorate programme.

The civilian and military staff at the RSME and the Clerk of Works electrical and mechanical students without whose support I could not have carried out my study and the senior officers at the RSME for permission to conduct the study.

The civilian and military staff at the School of Equipment Support (Army) for their support since joining the School as the Principal Lecturer.

The Abstract

The Army train their personnel using a system called 'The Systems Approach to Training' (SAT). The system develops the results of a job analysis into course training objectives that are used to define Army training courses. The outcome expected from the Army training system is individuals who are competent to perform the job they are trained for. The focus of this study is the electrical and mechanical military engineers trained by the Army to carry out engineering tasks throughout the world. They are trained in accordance with SAT but at the end of the courses they are also awarded a BTEC HND. The outcome of the courses is therefore to produce an HND qualified, competent military engineer. This study is concerned with how the students on the courses are assessed as competent and in what way this assessment is valid. There are different forms the validity could take, e.g. content, construct, and criterion-related, however the theory has moved away from individual forms to a unitary concept, with construct validity as the integrating force that binds it together. This study sets out to evaluate the nature of the validity of the course assessments in terms of this unitary concept. An evaluation of the validity of assessment requires more than an exploration of conceptions of validity, assessment content and assessment methods. As the assessment is used to assess military engineering competence this study is concerned with issues of competence-based assessment. Defining engineering competence as the construct presents two problem areas that threaten the validity of the assessment, domain specification and the use of the assessors' judgement. Specific theoretical criteria against which to assess the validity of the assessment must be defined. The criteria will focus on how the construct is represented in the assessment and the claims made for the assessment in terms of predicting occupational performance. This study will be an exploration into the nature of the validity of the assessment of military engineering competence. The results of this study will support the view that the assessment of competence and the validity of that assessment are complex issues, and my study will make a contribution towards the understanding of these issues for those concerned with the theory and practice of education.

CHAPTER 1

Introduction

The Origin of the Study

A major educational aim in the 1980's was to make assessment an integral part of the teaching and learning process: assessment was seen to be a tool of the curriculum. Implicit in this notion of assessment being a tool of the curriculum is that the assessment would be expected to be valid. This educational aim has resulted in almost two decades of discussion, development and re-conceptualization of the issues involved in assessment and its validity. As we begin the new millennium, emphasis remains on the importance of assessment and the need for educational institutions to think strategically. Any change in assessment that results from a strategic review must, surely, have expected positive consequences in terms of the impact on the teaching and learning processes. A good assessment programme will give teaching staff valuable feedback on how their students are progressing. It will act as a motivating and productive influence on the students by giving them feedback on how they are performing and on what they need to do for future development. Assessment can be seen as being connected to the process of teaching and learning in that it defines their nature, and what the students become through what they know and what they can do. The development of vocational education has increased significantly the importance of the issue of knowing *and* doing and a result of this has been the development and use of competence-based assessment.

During the 1980's and early 1990's, in parallel to the discussions on assessment, I was working as a chartered mechanical engineer. As a result of this I became aware of the need to develop the competence of graduate and apprentice engineers. This, coupled with the fact that I had qualified as a teacher in the late 1970's, lead to a developing interest in the education and training of engineers. My engineering and educational background merged when I took up the post in 1995 as the Senior Lecturer in Mechanical Engineering at the Royal School of Military Engineering. My

CHAPTER 1

Introduction

The Origin of the Study

A major educational aim in the 1980's was to make assessment an integral part of the teaching and learning process: assessment was seen to be a tool of the curriculum. Implicit in this notion of assessment being a tool of the curriculum is that the assessment would be expected to be valid. This educational aim has resulted in almost two decades of discussion, development and re-conceptualization of the issues involved in assessment and its validity. As we begin the new millennium, emphasis remains on the importance of assessment and the need for educational institutions to think strategically. Any change in assessment that results from a strategic review must, surely, have expected positive consequences in terms of the impact on the teaching and learning processes. A good assessment programme will give teaching staff valuable feedback on how their students are progressing. It will act as a motivating and productive influence on the students by giving them feedback on how they are performing and on what they need to do for future development. Assessment can be seen as being connected to the process of teaching and learning in that it defines their nature, and what the students become through what they know and what they can do. The development of vocational education has increased significantly the importance of the issue of knowing *and* doing and a result of this has been the development and use of competence-based assessment.

During the 1980's and early 1990's, in parallel to the discussions on assessment, I was working as a chartered mechanical engineer. As a result of this I became aware of the need to develop the competence of graduate and apprentice engineers. This, coupled with the fact that I had qualified as a teacher in the late 1970's, lead to a developing interest in the education and training of engineers. My engineering and educational background merged when I took up the post in 1995 as the Senior Lecturer in Mechanical Engineering at the Royal School of Military Engineering. My

CHAPTER 1

Introduction

The Origin of the Study

A major educational aim in the 1980's was to make assessment an integral part of the teaching and learning process: assessment was seen to be a tool of the curriculum. Implicit in this notion of assessment being a tool of the curriculum is that the assessment would be expected to be valid. This educational aim has resulted in almost two decades of discussion, development and re-conceptualization of the issues involved in assessment and its validity. As we begin the new millennium, emphasis remains on the importance of assessment and the need for educational institutions to think strategically. Any change in assessment that results from a strategic review must, surely, have expected positive consequences in terms of the impact on the teaching and learning processes. A good assessment programme will give teaching staff valuable feedback on how their students are progressing. It will act as a motivating and productive influence on the students by giving them feedback on how they are performing and on what they need to do for future development. Assessment can be seen as being connected to the process of teaching and learning in that it defines their nature, and what the students become through what they know and what they can do. The development of vocational education has increased significantly the importance of the issue of knowing *and* doing and a result of this has been the development and use of competence-based assessment.

During the 1980's and early 1990's, in parallel to the discussions on assessment, I was working as a chartered mechanical engineer. As a result of this I became aware of the need to develop the competence of graduate and apprentice engineers. This, coupled with the fact that I had qualified as a teacher in the late 1970's, lead to a developing interest in the education and training of engineers. My engineering and educational background merged when I took up the post in 1995 as the Senior Lecturer in Mechanical Engineering at the Royal School of Military Engineering. My

interest in vocational education thus developed from my work, and through studying with the Open University for an MA, into a growing realization of the need to investigate the issues surrounding the links between the curriculum, assessment, and the teaching and learning process. Much of the discussion about assessment has focused on its impact on the teaching and learning process and I felt that one area I wanted to investigate was that of the validity of assessment. Although I perceived the validity of assessment to be an extremely complex issue, this became the focus for my study. In carrying out my study I have tried to show how validity can be better understood as a concept and that it is a critical requirement of the assessment. My study became an inquiry into the validity of the assessment of engineering competence, an exploration of conceptions of validity, the definition of engineering competence, and how engineering competence is assessed.

The Context of the Study

The research was carried out at a military engineering training establishment, the Royal School of Military Engineering (RSME), and it focused on the Clerks of Works courses in electrical and mechanical engineering. These courses are accredited by BTEC and graduates are awarded an HND in Building Services Engineering. Hence, the Army and BTEC both input to the content of the courses. The electrical and the mechanical courses are each managed by a Senior Military Lecturer, a Warrant Officer who has served as a Clerk of Works engineer in the relevant discipline. The courses consist of three phases: a foundation phase, an engineering phase, and a civilian industrial attachment. Mathematics, science and computing are taught as the foundation phase modules and are followed by the subject-specific engineering modules, which are taught in the main engineering phase. The final phase is a six-week industrial work placement.

The foundation phase is taught and assessed by the Science & Computer Department. There are twelve lecturers within this department, ten civilians, two with a military background, and two military personnel. The

engineering phase is taught by the electrical and mechanical lecturers, with an equal mix of two civilian and two military lecturers for each discipline. The lecturers are either military personnel with practical engineering knowledge and experience, or civilian personnel who are qualified and experienced educators, some of whom also have practical engineering experience gained either from military service or industry. The military lecturers are posted on average every two years, whilst the civilians tend to supply the long-term continuity. The military lecturers supervise the industrial attachment phase, as directed by the Senior Military Lecturers. There are other engineering elements that require specialist lecturers and these are supplied from other departments, but controlled, requested and monitored by the two Senior Military Lecturers.

The students start the courses as Corporals and leave as Staff Sergeants, with an average starting age of twenty-six. Each course caters for a maximum of fifteen students, the actual number depending on the current military requirement and student availability. On passing the course they take up posts as military engineers, in roles primarily concerned with the technical design and installation of building services for the British Army. The course curriculum is designed around a job analysis carried out by the military Training Design Team (TDT). The job analysis process produces training objectives that are then combined with the BTEC HND unit specifications. The aim of the process is to ensure that the Army is supplied with competent engineers by identifying the specific training requirements related to the job of a Clerk of Works military engineer. Thus the courses are vocational in that they prepare a specific student group for a specific occupation. Students who pass the course assessment are thus deemed to be competent to fulfil this role.

Both the Army and BTEC recognize and discuss the need for some form of validity to be present in the assessment of the students' competence. However, the level of detail and their explanations of the concept of validity are very limited and there is a lack of clarity from both the Army and BTEC. The assessment of the students is designed and administered by the RSME

lecturers during the different phases of the courses thus the responsibility for ensuring the validity of the assessment rests with them.

The Aim of the Study

My study is concerned with exploring the issues involved in assessing military engineering students and, in particular, the nature of the validity of this assessment. The aim of my study is to test my hypothesis that the assessment of the students on the courses is valid and to establish the nature of its validity. The assessment could be valid because validity, in some form, is explicitly designed into the assessment, implicit in the assessment due to the nature of the assessment process, or both. This gives rise to the two sources of validity explored in my study, the explicit intention of those who design and administer the course assessment and the nature of the assessment process inherent in the culture of the organisation.

I do not believe anyone concerned with the course assessment process would argue that the assessment was invalid. Even if the assessment is seen to be valid simply because it measures *what it is supposed to measure* and this is stated as the requirement for validity, this simplistic view of validity still requires complex issues to be addressed. The assessment will only be valid if this requirement is met. To confirm that the assessment is indeed valid there are two questions that must be answered, what does the assessment actually measure and does it measure it *in such a way that any claims made for the results of the assessment can be justified.* So although I will identify the source of the validity of the assessment, these questions can only be answered from a study of the assessment process.

In my study, BTEC, the Army Training Organisation, and the lecturers all exert some influence on the validity of the assessment. To explore the way they influenced the validity there was a need to identify their understanding of the concepts of validity and which were applied when designing the course assessment. Whatever the concept of validity that was applied in practice, which I will show was the simplistic concept stated above, I wanted to establish a theoretical view of validity I would adopt for my

study. This is why I began my literature review with a discussion of the theoretical conception of validity and in so doing addressed the complex nature of validity as a concept. I will establish that the theoretical concept of validity, accepted by the leading theorists as being the most robust in terms of assessment, is centred on a unitary concept with a theoretical view of the construct at its heart. I will have established two concepts of validity, the concept adopted by the assessment designers and the unitary concept I adopted. I will show that the evaluation of the validity of the assessment in terms of the unitary concept encompasses all the issues in terms of the assessment being valid because it measures what is supposed to measure. An essential part of the evaluation will be to establish a set of theoretical criteria against which the validity of the course assessment can be measured. The criteria I will use will be based around two areas, how the construct is represented in the assessment and what claims are made for the assessment. By discussing the unitary concept of validity, identifying the criteria and then evaluating the assessment against them I will draw conclusions about the nature of the validity of the assessment.

Since the construct is the unifying force of the unitary concept, the evaluation of the assessment should begin by exploring what should be and what is actually being assessed. There are two elements to this part of my study, I will explore the *process of defining the construct*, how BTEC, the Army Training Organisation and the lecturers arrive at a definition, and what is *defined as the construct being assessed*. My study is concerned with the assessment of engineering students and hence a theoretical construct based on notions of engineering competence. I will show that defining competence is a complex issue and in terms of engineering competence, how it is defined determines what the engineer is trying to achieve and hence the knowledge, skills, attitudes and abilities the engineer must possess. My study will discuss how the theoretical construct is defined, using the theory from the available literature and in practice by those involved directly in engineering education at the RSME. I will explore the views of the Army, BTEC and the lecturers on how the construct engineering competence is defined, what judgements are being made, and by whom, on what the competent engineer must know and be able to do. In

this way I will show that *how* the theoretical construct is defined at a military establishment, where the engineers are being prepared specifically for jobs in the Field Army, does reflect the key elements and the problematic areas highlighted in the literature. However, establishing a theoretical model of the construct engineering competence does not mean that its assessment is necessarily valid. It is not how engineering competence is defined by the curriculum but how it is assessed that indicates the degree of fit with the theoretical construct and hence the construct validity of the assessment. I will not be concerned with what is taught in class but with what is measured in the assessment. An evaluation of what is being assessed and the methods used will show what is being defined and represented as the construct in the assessment, how the students are being assessed, and what judgements are being made about how engineering competence is inferred.

For those that employ the graduates, the result of the assessment infers that those that pass will be competent military engineers and my study is therefore concerned with issues of competence-based assessment. Competence cannot be measured directly but is inferred from performance assessed against a defined standard. Domain specification is an important feature of competence-based assessment that links with the discussion on the process of defining engineering competence as a construct and how it is represented in the assessment. Assessing student performance against a defined standard must involve the judgement of assessors. I will show that it is the lecturers, and in particular the military lecturers, that make judgements both about student performance and the standard it is judged against. In my study I want to look at how the lecturers make these judgements. I want to find out if they make their judgements based on defined assessment criteria or on their own model of what constitutes a competent military engineer. As a result of the judgements being made about student performance claims will be made about the assessment. The graduates will be considered competent to do the job of a military engineer and the assessment results interpreted as predicting current and future occupational performance. I will evaluate the claims made for the

assessment in predicting occupational performance and this will form an important element of my study.

The lecturers' evidence alone will not supply sufficient evidence to confirm validity in terms of the unitary concept, what is also required is an evaluation in terms of other evidence. The students are both participants and stakeholders in the assessment process and therefore their views, their perceptions, are also important. The students' views are important in supporting, or otherwise, any claims made by the lecturers in terms of the representation of the construct in the assessment and predicting occupational performance. Other views from serving Clerk of Works engineers and their employing officers about the course assessment, the competence of the students and the value of the input from the military lecturers will be presented. This will supply evidence for and against the claims made by the lecturers for the assessment. This data will come from research conducted by the RSME Training Design Team who surveyed the views of serving Clerk of Works engineers and their employing officers.

As a result of carrying out this process of validity inquiry I will be able draw conclusions about whether the assessment is indeed valid and if so, the source and the nature of its validity. This will satisfy the aim of my study in terms of the inquiry into the validity of the course assessment. However there is another important element of my study. I will also show the significance and the implications of my findings to those directly involved with the courses at the RSME and to those involved in the field of educational theory and practice. I will also draw conclusions about the process of conducting studies into assessment and its validity, highlighting the number and the complexity of the issues involved in that process.

CHAPTER 2

The Research Questions

In this chapter I will show how my research questions were developed and how they shaped the study. They supplied a route map showing my thought process and the way the research developed into this dissertation. I began with the hypothesis that the assessment of the students on the courses was, in some form, valid. This lead me to the decision during the early stages of my research to look initially at how the validity of assessment was conceptualized by the various stakeholders and participants in my study. I wanted to find out what concept of validity the lecturers used and what concept the Army and BTEC adhered to. In asking the first set of questions I wanted to establish two things, the source of the validity of the assessment and the importance given to the need to state explicitly the form of the required validity.

- In assessing the engineering students, how do BTEC and the Army interpret the meaning of validity?
- What types of validity do they expect to see as being present in the assessment?
- What guidance do they give to the lecturers assessing the courses on how they are to ensure the assessment is valid?
- How do the lecturers interpret the meaning of validity?
- What form of validity do the lecturers expect to be present in the assessment?

I set out to explore what judgements the lecturers made in deciding what form of validity should be present in the assessment, whether they followed the Army, BTEC or some other guidance. However this raised the question that if the data indicated that they did not follow any specific guidance or adhere to a particular concept, and validity in some form was not explicit in the assessment, what would this mean to my study? I considered the implication to be that it would render the guidance from both the Army and

BTEC irrelevant to the design and practice of the assessment of the courses, and I would be looking to identify what form of validity was implicit in the assessment. If, as I will show, the Army and BTEC literature was not referred to by the lecturers then my study needed to explore the implications of the lecturers lack of understanding or lack of recognition of the need to seek validity explicitly in the assessment. In other words, at this stage I would only be able to establish the source of the validity of the assessment and not the form of the validity present in the assessment.

In terms of the literature review these questions resulted in the need to look at the theoretical concepts of validity contained in the assessment literature. First of all I wanted to know how the stakeholders' conceptualization of validity related to the theoretical concepts. I also needed to establish my own theoretical perspective on the validity of assessment. I had to ensure that my study was conducted from a perspective that reflected current educational theory and would therefore result in a research programme of some value to those at the RSME and to those working in the wider educational context. From my exploration of the literature it emerged that the notion of the unitary concept, as developed by Samuel Messick (1989), with construct as the integrating force would be central to the perspective I adopted. So if validity of the course assessment was evaluated against the unitary concept then what would be the process my study should follow?

Initially this perspective made me return to my first set of questions. If the stakeholders did not seek to ensure the assessment was valid in terms of the unitary concept how should I proceed? Should I evaluate the assessment in terms of the theoretical validity criteria related to the unitary concept or should I look to substantiate the lecturers' claims for validity by looking for evidence that related specifically to their conceptual viewpoint. I wanted to go beyond an investigation based on the simplistic view of a valid assessment being one that measures what it is supposed to be measuring. I wanted to explore the implicit nature of the validity of the assessment and so I decided that my study must be an evaluation based on the unitary concept. I decided this evaluation process would in fact substantiate, or otherwise, any claims for validity made by the possible sources of validity by

identifying any form of validity implicit in the assessment. So to carry out the evaluation I needed to define the theoretical validity criteria against which to measure the validity of the assessment in terms of the unitary concept. This resulted in a need to answer the following questions.

- How would I define the theoretical criteria against which I would assess the validity of the course assessment?
- What would be the theoretical criteria I would use?
- What would be the impact of using these criteria on my research questions?

From the assessment literature I was able to identify and discuss Messick's (1989) unitary concept and use this concept as the main framework on which to base the theoretical validity criteria. It became clear that I would need to focus on issues concerned with the construct engineering competence and the claims made for the assessment in terms of how the results of the assessment were being interpreted and used. As my study was concerned with assessing engineering competence, I was exploring the validity of competence-based assessment and so issues of domain specification, assessors' judgement, and predicting occupational performance became important.

The need to establish that the assessment had construct validity was evident from the assessment literature. I needed therefore to explore how the construct engineering competence was defined from the literature and in practice at the RSME. Since the courses had an Army and a BTEC requirement, the process of defining the construct was important. I needed to explore the degree of similarity between the theory and the practice. So in terms of generating data, these issues resulted in a further set of research questions.

- How do BTEC and the Army define the construct that is to be assessed?
- Does a conflict exist between the BTEC and the Army view of the construct that is to be assessed?

- Do the lecturers have a clear definition of the construct they are to assess?
- Do the lecturers use their own judgement in deciding on the definition of the construct they assess?

If there was a conflict between the Army and BTEC systems, it was the nature of the conflict and how it was resolved that was important. How it was resolved would show what judgements were being made, and by whom, about what constituted the construct engineering competence. However, all this would reveal was that a model of the construct existed and the process that defined it. I considered there were two ways of establishing the definition of military engineering competence, from what was taught or from what was assessed. I decided that there were two reasons why I would not study what was taught, the first was a logistics problem in that I felt this was a study in itself. The second reason was that I would still have to look at what was assessed and would even then only be able to discuss the assessment in terms of content validity. It was evident to me that my next step would have to be to establish how the construct was represented in the assessment and to address the following questions.

- Who designs and administers the assessment?
- What methods of assessment are used?
- Are there any laid down requirements in terms of the number and type of assessment methods to be used on the courses?
- Is the construct adequately represented in the assessment?
- Are the students' responses likely to be different for some reason that is irrelevant to the construct being measured?
- What claims are made for the assessment?

The answers to these questions will build a picture of what is being assessed and supply data that will be analysed in terms of construct validity. All the questions so far have dealt predominantly with issues concerning the process by which engineering competence was being defined, how it was represented in the assessment and the use of the lecturers' judgement throughout the assessment process. There was a need to explore evidence

from other sources in order to establish triangulation. What was required was triangulation of the evidence in terms of the use of the lecturers' judgement in defining the construct, representing it in the assessment, and for the claims being made with regard to the interpretation and use of the results of the assessment.

- How and what do the students believe is being assessed?
- Do the students' respond differently when assessed for any reason that is irrelevant to the construct being measured?
- How do experienced Clerk of Works engineers and their employing officers view the training and the students who pass the assessment?

The answers to these questions will supply other evidence about the validity of the assessment. The evidence will not just be about the assessment of the construct but also about whether the assessment does what it claims to do.

The answers to all these research questions will supply data that addresses the issues raised in my study of assessment and its validity. The way my study is carried out depends on the answers to three questions.

- What sort of study will I conduct?
- How will I collect the data?
- How will I analyse it?

I decided that I was going to carry out a qualitative case study. My aim was that the data I would be collecting from interviews, questionnaires and documentation would be analysed using qualitative methods. I wanted to conduct a study that explored and identified issues and problems concerned with validity of the assessment of engineering competence. From my knowledge of the courses and the student population I concluded that detailed statistical analysis of the data would not be appropriate, however some of the data generated was quantitative and resulted in the use of simple percentages and numerical tables being used as part of the data analysis.

The study will conclude with answers to two final questions.

- In what way is the assessment of engineering competence at the RSME valid?
- What are the implications of my study to educational theory and practice?

Only when I have addressed all the previous questions will I be able to support, or challenge my hypothesis that the assessment is indeed valid and state the nature of its validity. My study will provide an insight into the educational practice at the RSME and in answer to the final question, I will discuss the significance of my study to people concerned with the theory and practice of education, particularly in relation to the field of occupational education.

CHAPTER 3

The Literature Review

Introduction

In order to establish the validity of assessment I must have a clear understanding of the concepts of validity developed from the assessment literature. Validity, at its simplest, is seen as being the extent to which an assessment actually measures what it is intended to measure (Ashcroft & Palacio, 1996; Cotton, 1995; Futcher, 1989). However, it is accepted that any attempt to develop a deeper explanation of the meaning of the term validity is problematic (Ashcroft & Palacio, 1996; Cotton, 1995) and it is widely acknowledged that it is a very complex concept (Dary-Erwin, 1995; Gipps, 1995; Linn & Baker, 1996; Wood, 1991). It is stated generally that assessment should be both reliable and valid (Tolley, 1989; Futcher, 1989; Nuttall, 1989; Gipps, 1995; Cotton, 1995; Linn & Baker, 1996). There is evidence to suggest that reliability has been emphasized at the expense of validity (Ashcroft & Palacio, 1996; Brown & Knight, 1994; Gipps, 1995; Wood, 1991) and my study forms part of the movement to attempt to redress the balance. It is my intention to explore the complexity of the issues surrounding validity of assessment and to establish a framework upon which to develop my study. Central to this framework is the notion that validity should be viewed as a unitary concept with construct as the integrating force. Within this framework I must discuss what makes an assessment valid and establish the theoretical criteria against which I will evaluate the validity of the assessment of the courses in my study.

In the institution that is the focus of my study the outcome of the assessment process is the production of a competent military engineer. This is an individual who within days of completing the course could be sent anywhere in the world and be expected to deal with complex military engineering problems. Engineering competence must therefore be the theoretical construct that is assessed in an appropriate manner such that those individuals who pass the assessment can be considered competent. Hence the second part of this chapter is concerned with competence-based

assessment, the notion of competence and the definition of engineering competence as a theoretical construct. The chapter will be completed by a discussion on the use of assessors' judgement and the process of aggregation. If the results of the assessment are used to employ individuals in jobs as competent military engineers then all the above issues need to be addressed in order to make any claims about the validity of the course assessment.

Conceptions of Validity

Discussing Concepts of Validity

My study is focused on the issue of validity of assessment and my first objective must be to establish what is understood by the term validity and to identify the perspective I will take throughout the study. Four main categories of validity first surfaced from a paper by Cronbach and Meehl (1955): *content*, *construct*, *concurrent* and *predictive* validity. In some texts concurrent and predictive validity have been combined to form *criterion-related* validity and it is suggested that this is so because they both relate to predicting performance on some criterion, either at the same time, or at some time in the future (Gipps, 1995; Messick, 1989). Combining these two, Wood (1991) concludes, gives rise to the three *types* of validity, content, construct and criterion-related, referred to by Guion (1980) as the *trinitarian* view. However, on exploring texts written on the theory and practice of assessment other types of validity appear, *face* and *consequential* are two such examples (Ashcroft & Palacio, 1996; Benett, 1993; Brown *et al*, 1997; Cotton, 1995; Sambell *et al*, 1997). Face validity is referred to in relation to work-based assessment and is therefore worth discussion, although it falls outside the more traditionalist views discussed above it is referred to in relation to the validity of vocational assessment. Consequential validity will be dealt with later in the discussion of the unitary concept of validity.

What follows are the general definitions of four of the specific types of validity mentioned above: content, construct, criterion-related, and face. The

first three will be shown later as being subsumed by Messick's unitary concept of validity. The discussion on conceptions of validity is centred on the assessment of educational constructs that are essentially curriculum-based however it is recognized that conceptions of validity do also apply to the assessment of other constructs, e.g. psychological and behavioural.

Content Validity. Content validity is concerned with the coverage of appropriate and necessary content. It is based on expert or professional judgements about the relevance of the content to that of a particular behavioural domain of interest and about the way the item or task content is represented in covering that domain (Gipps, 1995; Messick, 1989). The sample of tasks must be a sufficient representation of the domain that is, itself, clearly delineated. However the size and complexity of the domain may be such that delineation becomes problematic (Benett, 1993).

Construct Validity. Construct validity relates to whether the test is an adequate measure of the construct. A clear and concise definition of the construct is of paramount importance and to highlight this reading has been used as an example of a construct:

> ...a full definition of reading as a construct would include not only reading aloud, but also reading comprehension, accuracy, and enjoyment of reading. (Gipps, 1995, p. 58)

This definition reflects the combination of curriculum content, cognitive processes and behavioural outcomes that Haertel (1985) used to describe the construct. It is defining the construct that make studies of construct validity the most difficult to conduct, a view supported by Wood (1991). In my study the construct is military engineering competence and this means that a full definition of this construct would include all the elements, or competencies, that go to make up engineering competence. In other words, the definition must include the underpinning knowledge, the skills, the attitudes and the abilities that I will discuss later in this chapter.

Criterion-Related Validity. Criterion-related validity refers to the extent to which an individual's performance on a criterion measure can be estimated from the individual's performance on the assessment procedure being validated (Salvia & Ysseldyke, 1998). It relates to the combination of *concurrent* and *predictive* validity. Concurrent validity is concerned with how accurately an individual's current performance estimates performance on the criterion measure at the same time. In concurrent validity the assessor has to identify the trait which needs to be assessed so that the correct assessment method can be chosen that should give the same results substantially as another assessment of the same trait (Cotton, 1995; Gipps, 1995). Predictive validity relates to whether the test predicts accurately future rather than current performance and is useful for specific purposes such as assessing a set of well-defined skills which are used in a profession (Brown *et al*, 1997). Although paper and pencil tests may provide good predictions of ability the available evidence suggests that the predictive validity of simulated job-specific tasks tends to be higher when compared against criteria of future occupational performance (Brown *et al*, 1997; Nuttall, 1989; Wolf, 1995; Wood, 1991). This does not mean that there is not a place for pen and paper tests in the assessment of occupational performance and that such tests would not have predictive validity. For example in the medical profession pen and paper tests are used to predict an individual's ability to choose a suitable method of patient care whereas simulated work-based tasks are used to predict an individual's ability to use patient consultation skills (Brown *et al*, 1997). The difficulty with predicting future performance is that there is no evidence to say how well those that fail would have performed subsequently in the workplace (Gipps, 1995; Wood, 1991). I will return to the issues of criterion-related validity in the next section on the unitary concept of validity and when I discuss assessing competence later in the chapter.

Face Validity. This is concerned with the *look* of the assessment. Does it appear to be assessing what it is supposed to be assessing? A common feature of vocational assessment is a reliance on face validity, whether something looks right to experts in the field, i.e. that the assessment appears to measure what it is supposed to measure (Gipps, 1995; Wolf, 1996). The

problem here is that the assessment may look right but is likely to be *biased, unreliable, partial in coverage, or weighted in ways which do not correspond to the demands of the occupation* (Wolf, 1996, p. 212). However the importance of face validity is in the credibility it gives to the assessment in the eyes of the students.

In texts on assessment practice in education, e.g. Brown *et al* (1997); Cotton (1995); Cox & Harper (2000), validity is generally covered in a few pages of text containing very basic definitions of the types discussed above. Definitions are given which vary surprisingly between texts and there is a lack of explanation on how to ensure that any one of the types of validity would be present in the devised assessment. As published texts concerned with assessing student learning dedicate less than 1% of their content to the issue of validity of assessment and do not discuss how to design in validity, what can reasonably be expected of assessment designers in practice such as the lecturers in my study? What will be their depth of understanding of the concepts and how much emphasis do they put on the need to ensure the validity of their assessment? I will show that in my study the assessors are not familiar with the concepts discussed above and they simply perceive that the course assessment is valid because it measures what it is supposed to measure. This means that in order to explore the nature of the validity of the assessment I must adopt a theoretical concept of validity and describe how to evaluate the validity of the course assessment in terms of this concept. The next section deals with these issues and will show why I have adopted the unitary concept of validity and the theoretical criteria against which the validity of the assessment will be evaluated.

The Unitary Concept of Validity

The leading theorists, Linn, Guion, Landy, Anastasi, Cronbach and Messick, have all shifted away from the idea of separate types of validity towards the view of a unitary concept and they warn that excessive devotion to one distinct form, content, construct or criterion, is mistaken (Wood, 1991). It is argued that Cronbach has to be listened to as he is indeed challenging his own earlier ideas and that this, therefore, is the way forward for validity

enquiry (Gipps, 1995; Wood, 1991). Further it is argued that drawing distinctions between content, construct and criterion-related validity has *proved to be especially insidious because it implied that there were testing purposes for which one or another type of validity was sufficient* (Messick, 1989, p. 6). The theoretical discussion of validity centres on two main protagonists, Cronbach and Messick, although Messick is thought to be the main authority (Dary-Erwin, 1995; Gipps, 1995; Linn & Baker, 1996; Moss, 1992; Nuttall, 1989; Wood, 1991). Gipps (1995) concludes that Messick is the key writer and that his development of the unitary concept of validity takes the theory into a realm where both educational and social implications of assessment are addressed. It is the nature of this unitary concept that I will now discuss in more detail.

Messick (1989) argues that although considerations of content relevance and representativeness should and do influence the nature of score inferences when supported by other evidence, content validity does not qualify as validity at all. Content validity provides judgemental evidence on what is represented in the assessment and is thus making a statement about the assessment itself and not what can be inferred from it. The problem of irrelevant test variance contributes to *the ultimate frailty of classical content validation* since *expert* judgements which are made about the relevance of test content to the domain are indeed fallible and *may imperfectly apprehend domain structure or inadequately represent test structure, or both* (*ibid*, p. 7). Messick (1989) states that the way around this problem is to evaluate expert judgement on the basis of other evidence about the behavioural domain and the structure of assessment results i.e. construct-related evidence. The implication of this for my study is that although I will be exploring the judgements made by the lecturers I must also seek out other evidence that substantiates or refutes the claims being made by the experts, the lecturers.

Criterion-related validity was shown earlier to be concerned with the measurement of criteria in order to make predictive evaluations about individual performance and there must be as many criterion-related validities as there are criterion measures and settings within which they

apply. This introduces the first problem, generalization of the criterion correlation and the second problem is that of irrelevant variance. Thus criterion validity is examined in a similar manner to content validity and Messick argues that

> potentially deficient and contaminated criterion measures cannot serve as the unequivocal standards for validating tests, as is intrinsic in the criterion-oriented approach to validation. (Messick, 1989, p. 7)

The way around this, Messick argues, is to evaluate the criterion measures and the assessment in relation to construct theories of the criterion domain.

Construct validity is based on *an integration of any evidence that bears on the interpretation or meaning of the test scores* (Messick, 1989, p. 7). The convergent and discriminate nature of construct-related evidence is seen as critical in the delineation of content domains and in the conceptualization and measurement of applied criteria providing a rational basis for evaluating both content and criterion-related validity. If all validity evidence contributes to the empirical grounding or trustworthiness of the interpretation of results, and the validity of this interpretation provides the main rationale upon which inferences and actions are drawn from assessment results, then Messick (1989) argues all validation is construct validation.

The argument presented gives rise to two major threats to validity (Gipps, 1995; Messick, 1989). The first, *construct under-representation*, is defined as being when the assessment is too narrow and has failed to include important dimensions of the construct. The second, *construct-irrelevant variance* occurs when the assessment contains *excess reliable variance*, meaning that some respondents will find the assessment items easier or more difficult in a way that is irrelevant to the interpreted construct. Although almost any form of information about an assessment contributes to an understanding of its construct validity, this contribution is extremely

enhanced if *the degree of fit of the information with the theoretical rationale underlying score interpretation is explicitly evaluated* (Messick, 1989, p. 7).

The unitary concept of validity integrates considerations of content and criterion-related validity into a construct framework however there is a third consideration that also needs to be incorporated in the concept of validity, social consequences. Messick (1975) was the first theorist to suggest that there were two questions that need to be asked whenever a decision about test use is made.

> First, is the test any good as a measure of the characteristic it is interpreted to measure? Second, should the test be used for this purpose? The first question is a technical and scientific one and may be answered by appraising evidence bearing on the test's psychometric properties, especially construct validity. The second question is an ethical one, and its answer requires an evaluation of the potential consequences of the testing in terms of social values. (Messick, 1975, p. 962)

This statement already contained the elements that were later developed into the progressive matrix that defined Messick's facets of validity and most importantly drew attention to the incorporation of social consequences. Messick (1989) defined two interconnected facets of the unitary concept of validity, each of which had two parts. The first was the source of justification of the testing divided into an appraisal of either evidence supportive of score meaning or of consequences contributing to score valuation. The second facet was the function or outcome of the testing divided into either interpretation or use. From these facets he constructed a fourfold classification matrix (Messick, 1989, p. 10, copyrighted to the American Educational Research Association). The matrix has construct validity appearing in every cell and this is *fitting because construct validity is the integrating force that unifies validity issues into a unitary concept* (Messick, 1989, p. 10). The matrix is seen as progressive because it moves from appraisal of evidence for the construct interpretation, evidence

supportive of test use, the value consequences of score interpretation and, finally, the social consequences of test use (*ibid*). In terms of social consequence Messick argues that

> In general, the best protection against adverse social consequences as threats to valid score interpretation and use is to minimize in the measurement process any potential sources of test invalidity, especially construct under representation and construct-irrelevant variance in the test. Thus, the watchword for educational and psychological measurement is to maximize empirically grounded interpretability and minimize construct irrelevancy in the test scores. (Messick, 1989, p. 11)

Haertel (1991) adopts a similar perspective to that of Messick (1989) as to the relative importance of social consequences, that adverse impact is not sufficient to undermine validity. Messick (1989, p. 6) discussed six basic sources of validity evidence under the heading *Ways of Configuring Validity Evidence*. In order to collect evidence and arguments to discount the two major threats identified above these headings were further developed into a set of questions that he considered must be addressed in the design and development of any assessment. The questions listed are as follows:

- Are we looking at the right things in the right balance?
- Has anything important been left out?
- Does our way of looking introduce sources of invalidity or irrelevant variance that bias the scores or judgements?
- Does our way of scoring reflect the manner in which domain processes combine to produce effects and is our score structure consistent with the structure of the domain about which inferences are to be drawn or predictions made?
- What evidence is there that our scores mean what we interpret them to mean, in particular, as reflections of knowledge and skill having plausible implications for educational action relative to personal or group standards?

- Are there plausible rival interpretations of score meaning or alternative implications for action, and, if so, by what evidence and arguments are they discounted?
- Are the judgements or scores reliable and are their properties and relationships generalizable across the contents and contexts of use as well as across pertinent population groups?
- Do the scores have utility for the proposed purposes in the applied settings?
- Are the scores applied fairly for these purposes?
- Are the short- and long-term consequences of score interpretation and use supportive of the general testing aims and are there any adverse side effects?

(Messick, 1992, p. 3)

Messick's argument and questions are not seen as being the perfect solution to ensuring the validity of assessment (Gipps, 1995; Moss, 1992; Shepard, 1993; Tittle, 1989). Moss (1992) presented a critique of the works of a number of authors, including Messick, that provided a

> description and synthesis of various sets of categories that have been used by Messick (1980, 1989) and Cronbach (1988, 1989) writing in the context of assessment in general, and by Frederiksen and Collins (1989), Haertel (1990, 1991), and Linn, Baker and Dunbar (1991) writing in the context of performance assessment in particular.
>
> (Moss, 1992, p.231)

That Messick considered his set of questions should be applied to *any* assessment and that both he and Cronbach were concerned with the validity of assessment in general indicates clearly that conceptions of validity are not confined to the assessment of educational constructs. The last three sets of authors listed above all built on the work of Cronbach and Messick and pointed out that the practice of validity research had not done justice to the views Cronbach and Messick articulated. Whilst recognizing that conceptions of validity can be equally well applied in other contexts, for

example psychometric testing, the work of the three sets of authors is important to my study because of the educational context within which they deal with conceptions of validity. Moss (1992), followed by Gipps (1995), critiqued the additional criteria or advice they proposed that addressed concerns about meaningfulness and directness raised by those involved in competence-based assessment. All the authors argued that the consequences of assessment-based interpretations and actions needed to be considered, particularly in terms of teaching and learning. The notion of systematic validity developed by Frederiksen and Collins (1989) further emphasized the consequences of assessment practices (Gipps, 1995; Hickey *et al*, 1999; Linn *et al*, 1991; Moss, 1992; Sambell *et al*, 1997). A systematically valid assessment is one that *induces in the educational system curricular and instructional changes that foster the development of the cognitive skills that the test is designed to measure* (Frederiksen and Collins, 1989, p. 27). The main characteristics that contribute to systematic validity are the directness of the cognitive assessment and the degree of judgement required in assigning a score. The advantage of direct assessment is that instruction that improves the assessment score will also result in improved performance on the extended task and on the expression of the cognitive skill within the context of the task. Thus teaching to the test becomes teaching to the domain (Frederiksen and Collins, 1989). Gipps (1995, p. 102) argues that *given the promises made for performance assessment the concept of systematic validity is relevant.*

Linn, Baker and Dunbar (1991) produced criteria for validity evaluation that sit under the umbrella of construct validity. These are cognitive complexity of the task with its analysis going beyond face validity, content quality and coverage, the tasks must themselves be worth the time and effort of the students, and subject matter experts must be involved in assessment design. The last point is concerned particularly with the impact of assessment on teaching, if there are gaps in the content covered by the assessment then *teachers and students are likely to under-emphasize those parts of the content domain that are excluded from the assessment* (Linn *et al*, 1991, p. 20).

The authors critiqued by Moss (1992) and Gipps (1995) wrote their articles for professional assessment developers and evaluators and as a result *students' interpretations were seen as secondary* (Moss, 1992, p. 251). Sambell *et al* (1997) focused on the effects of assessment on the teaching and learning context and social consequences of the use of assessment information by *drawing on qualitative data to illuminate the impacts of assessment practices on student perceptions of learning and on their learning behaviour* (*ibid*, p. 349). Another study that was concerned with consequential validity and students' interpretations was carried out by Hickey *et al* (1999) who adopted Shepard's (1993) alternative conceptualization of Messicks' model.

Although Shepard (1993) agreed with Messick on the scope and range of validity evaluation, she argued that his framework does not help to identify which validity questions were essential to support the use of a test. She produced what she considered to be a straightforward means to prioritize validity questions and suggested that

> validity evaluations be organised in response to the question: What does the testing practice claim to do? Additional questions are implied: What are the arguments for and against the intended aims of the test? and What does the test do in the system other than what it claims, for good or bad? All of Messick's issues should be sorted through at once, with consequences as equal contenders alongside domain representativeness as candidates for what must be assessed in order to defend test use.
>
> (*ibid*, p. 429)

So where does this leave me in terms of my own study? In discussing the unitary concept of validity I identified the two major threats as construct under-representation and construct irrelevant variance. I must show whether there is sufficient evidence to discount them as threats to the construct validity of the course assessment. To do this Messick, as Gipps (1995) argued, would have me address validation requirements of extensive

proportions as indicated earlier by his list of questions. There appears to be consensus amongst Linn, Baker and Dunbar (1991), Frederiksen and Collins (1989), and Shepard (1993), that consequential categories are as equally important as evidential categories. Also the study by Hickey, Wolfe and Kindfield (1999) would suggest that using just Messick's model might prove to be limiting. I am presented with the difficulty of deciding exactly what theoretical criteria I will use in my study to evaluate the validity of the assessment. I will look at what the assessment process claims to do in terms of measuring the construct, predicting occupational performance, and impacting on the teaching and learning process. I will then be able to draw conclusions about the validity of the assessment and also about the significance of my study to others involved in the theory and practice of education. The criteria I will use will be that:

- The construct is represented in the assessment in sufficient breadth and depth.
- Important elements of the construct have not been omitted from the assessment.
- The assessment does not introduce sources of invalidity or irrelevant variance that bias the scores or judgements.
- The way the assessment is marked reflects the manner in which domain processes combine to produce effects and is consistent with the structure of the domain about which inferences are to be drawn or predictions made.
- The results of the assessment can be interpreted to mean students that pass have demonstrated the required level of performance to be considered competent military engineers.
- The results of the assessment can be used to place graduates of the courses into jobs in the Field Army.
- A consequence of interpretation and use of the results of the assessment is that graduates go on to perform successfully in their occupational role.
- A consequence of interpretation and use of the results of the assessment is that there is a positive impact on the teaching and learning process.

I have now produced the theoretical criteria I will use as a basis to evaluate the validity of the course assessment. I have chosen the above criteria because my study is an exploration into the nature of the validity of the assessment of competence that relates to a defined occupational role, the military Clerk of Works engineer. It is my intention to explore the assessment process using these criteria as a framework for my exploration. I have made this decision based on the nature of the study I want to conduct. Chapter 4 deals with the issues of the methodology and methods in more detail. I will now focus on a discussion of the issues involved in the assessment of occupational competence.

Assessing Engineering Competence

I am concerned with issues of competence-based assessment and defining engineering competence as a theoretical construct. This section will be a discussion of the methodology of competence-based assessment and the development of the theoretical construct, engineering competence. At the end of this section I will discuss the use of assessors' judgement within the context of competence-based assessment.

The Relationship between Criterion-Referenced and Competence-Based Assessment

From the early 1960's, when Glaser (1963) made reference to criterion-referenced assessment, and 1991 some 900 papers have been published on the topic and many definitions of criterion-referenced assessment have been proposed (Hambleton *et al*, 1991). Criterion-referenced assessment has become a familiar term to those involved in education, it is associated with assessing performance against clearly specified outcomes, with assessing for achievement rather than against pass marks or norms, and hence recognized as the alternative to norm-referenced assessment. Glaser (1963) defined criterion-referenced assessment specifically in terms of its difference from norm-referenced assessment. The most widely accepted definition of a criterion-referenced assessment is an assessment that is *used to ascertain an individual's status with respect to a well-defined behavioral domain*

(Popham, 1978, p. 93). Competence-based assessment is considered to be a specialized development of criterion-referenced assessment, starting from and remaining entirely focused on outcomes or criteria (Jessup, 1991; Hager *et al*, 1994; Wolf, 1995, 1996). The difference between the criterion-referenced and competence-based systems is seen to be that criterion-referenced assessment is associated generally with mainstream education, criteria defined by the curriculum and paper-and-pencil tests, whereas competence-based assessment is essentially non-academic, focused on a notion of competence and defined in a vocational context (Wolf, 1995).

In my study, the aim of the courses, both mechanical and electrical, is to produce a competent military engineer with a BTEC HND qualification. The courses are assessed in terms of vocational and academic goals that involve both work-based and college-based learning. The relationship between how each of these elements is assessed is an area that should be explored (Usher & Bryant, 1987; Benett *et* al, 1989). The term vocational is traditionally linked to manual skills, derived from a craft tradition, and referenced to outside, occupational activity. There has been a significant shift in the way vocational assessment is viewed. It is no longer related only to craft skills but also to the professions and higher education. Professional bodies are now seen to have a key role in the development of students who are able to perform competently in their chosen professions (Boud, 1995; Hillier, 1997; Wolf, 1995). The elaboration and development of what is essentially a work sample approach, i.e. carrying out assessment of a work related task centred on a notion of competence, is seen as the single most important development in vocational assessment (Wolf, 1996).

Defining and Assessing Competence

There are many different concepts of competence and it is widely accepted that how it is conceived makes a significant difference to the assessment process (Benett, 1993; Hager, Gonczi & Athanasou, 1994; Hillier, 1997; Wolf, 1995). Benett (1993) states that the claim is made that competence is a *hypothetical construct* relating to what individuals are

> theoretically able to do and this ability is judged by the performance of what they actually do in particular circumstances. (Benett, 1993, p. 86)

Hager *et al* (1994) promote an integrated conception of competence in which competence is conceptualized in terms of

> knowledge, abilities, skills and attitudes displayed in the context of a carefully chosen set of realistic professional tasks which are of an appropriate level of generality. (*ibid*, p. 4)

Key tasks or elements that are central to the professional practice are selected and the main attributes required for competent performance of these tasks are then identified. Hager *et al* (1994) argue that experience has shown that the integrated concept captures the *holistic richness of professional practice* and one reason given for this is that

> competence is a construct that is inferred from performance of relatively complex and demanding professional activities. (*ibid*, p. 5)

Boud (1995) supports the above view and believes that moves are being made towards a holistic conception of competence that leads to a need to look at the *impact of the total package of learning and assessment and not simply at fragments of assessment* (Boud, 1995, p. 38).

The emphasis should not be on the practice alone. There must also be reference to the underpinning knowledge that is seen as inherent in the practice but there is a difficulty here in that much of this underpinning knowledge is not easily identified in practice because it is tacit (Eraut, 1994; Hager *et al*, 1994; Hillier, 1997). This is similar to the view expressed by Messick (1984) who makes the point that the structure of knowledge and abilities are embraced by the term competence. Messick (*ibid*) also states that there is a difference between competence and performance, the former

being about what an individual knows and does under ideal circumstances and the latter what they actually do under existing circumstances.

Competence is inferred from performance and this is so because the latter is observable whereas the former is not. Competence combines the possession of relevant attributes i.e. knowledge, skills and attitudes, with the ability to use them to perform certain tasks, or in certain roles or situations (Gonzi *et al*, 1993). In other words possessing knowledge, skills and attitudes alone does not constitute competence nor does just the performance of a series of tasks. Competence comes from the ability to use the former appropriately in the successful completion of the latter. Assessment of competence involves a judgement about performance. Someone assessed as competent is seen as a professional that can be trusted with a degree of responsibility in those areas within the range of their competence (Eraut, 1994). This is not to suggest that there are different grades of competence but that there is a level of task, a degree of difficulty, at which people cease to be considered competent. Judgements are made in terms of the distinction between novice and expert and Eraut (1994) suggests that there are two dimensions to deal with when defining competence in this way, scope and quality. *Scope* is defined by the range of roles, tasks and situations for which competence is established or inferred and *quality* is defined by the degree to which the individual can perform in these roles, tasks and situations (Eraut, 1994, p. 167). This notion of competence is based on a model of progression and lifelong professional learning that

> contrasts with those definitions of competence adopted by most competency-based systems of training and education, which assume a binary scale by confining assessment decisions to judging whether a candidate is competent or not yet competent.
>
> (Eraut, 1994, p. 215)

I agree that Eraut's model allows for the quality of professional work to be developed as a lifelong process and that progression beyond qualification is important. Expressed in simpler terms students graduating from the courses

are not experts but they have demonstrated the capability to perform, through some form of assessment, to the standard required of them at that time in their development. In my study, they would be considered by the assessors to be competent to start work as a Clerk of Works engineer, the results of the assessment predicts this and that the graduates have the potential to develop professionally as they continue with their careers.

Since the validity of competence-based assessment has an *external referent that derives from the world of employment* (Wolf, 1996, p. 212), and is judged by what the individual does after the assessment, then the focus should be placed on the key purposes of the occupation and the functions this involves. What must occur is some form of functional analysis of the outcomes of the occupation that focuses on their purpose and function (Eraut, 1994; Wolf, 1995, 1996). From this analysis the performance criteria will be developed that are central to the competence-based system and *reflect the critical aspects of performance - all those qualities which are essential to competent performance* (Fletcher, 1991, p. 169). By focusing on the general functions of the job the underlying, general requirements on which assessment should concentrate will be far more evident than through the use of a job analysis process that concentrates solely on current tasks and routines (McMahon & Carter, 1990; Wolf, 1996).

As stated earlier, for both criterion-referenced and competence-based assessment systems the behavioural domain needs to be specified clearly and unambiguously (Gipps, 1995; Hambelton *et al*, 1991; Shepard, 1991; Wolf, 1995, 1996). If the assessment is used to make statements about an individual's behaviour or competence in a particular area then there must be a one-to-one match between the assessment and the domain. This is where the problem lies for both assessment systems, the harder the attempt to define the domain being assessed, the *narrower and narrower the domain itself becomes, without, in fact, becoming fully transparent* (Wolf, 1995, p. 55). Domain specification is a problem area that has been widely recognized (Benett, 1993; Eraut, 1994; Hager *et al*, 1994; Hillier, 1997; Wolf, 1995). Wolf emphasizes the problem when she argues that in practice, the greater the requirement for clarity, the more tightly *the domains or outcomes or*

criteria tend to be defined, the more narrow and more numerous they become (Wolf, 1996, p. 222). This is a problem referred to as the *Catch 22 situation for criterion-referenced studies* (Gipps, 1995, p. 87). As this problem applies equally well to competence-based assessment, this directs Wolf (1996) to the conclusion that the validity issue is not as easily solved as those advocates of competence-based assessment would lead us to believe. She argues that

> analysing competencies is not enough if one cannot actually assess them directly, either because the outcome statements are essentially non-assessable or because of the simple volume of discrete tasks generated. (Wolf, 1996, p. 226)

The University of Houston example described by Eraut (1994) and concerned with competence-based teacher education highlights similar issues and he concludes that the

> main difficulties were the sheer number of objectives to be assessed, with consequent lack of proper attention to each; and the lack of any valid theoretical construct for combining or prioritizing assessment evidence.
>
> (Eraut, 1994, p. 173)

It has been suggested that one way to get around the problems associated with the assessment of objectives or outcomes could be to include periods of experience that supplement or replace direct assessment (Wolf, 1996). In terms of occupational performance this would be arguably the closest form of experience to vocational reality. Occupational performance is much more likely though to be assessed using academic measures or competence-based work sample methods. There is evidence that the correlation between academic measures and occupational performance is generally low (Ghiselli, 1996; Samson *et al*, 1984; Wood, 1991). One point that arose from a study by Williams and Boreham (1971) that is worth noting in relation to my study was that performance in mathematics, physics and

engineering drawing examinations could provide reasonable predictions of success in terms of passing engineering courses. Wolf (1995) critiques these studies and sums up the available evidence by stating

> Nonetheless, even allowing for these relationships, the case *against* [*my italics*] academic-style assessment as a way of predicting later vocational success is fairly strong. (Wolf 1995, p. 48)

Evidence from the competence-based approach suggests that the correlation between simulated work-based measures and occupational performance is much higher than for academic measures (Asher & Sciarrino, 1974; Robertson & Downs, 1979; Robertson & Kandola, 1982). Longitudinal studies of occupational success, for example in the Armed forces with their centralized training and assessment centres, are another source of evidence (Gardner & Williams, 1973; Feltham, 1988 a, b). Again Wolf (1995) has critiqued the studies and sums up the evidence by stating

> Overall, then, the evidence provides considerable support at a theoretical level for the competence-based approach. Arguing from first principles, we can conclude that faithful simulation and sampling of the behaviour of interest should provide us with the most valid form of assessment. (Wolf 1995, p. 52)

My study is concerned with assessment of occupational performance and for an academic qualification. The methods of assessment used, and the interpretation and use of the results of the assessment, will need to be evaluated in the context of the above discussion.

So far the discussion has been concerned with representing and assessing the construct competence. As I am concerned with competence in terms of engineering, the next section will be focused on how engineering competence can be defined as a theoretical construct.

Defining the Construct Engineering Competence

The process of defining the theoretical construct engineering competence must focus on what competent engineers need to know and do. Defining engineering competence as the theoretical construct is not in itself an easy task, as Vincenti (1993, p. 15) states *the epistemology of engineering knowledge is obviously at its beginnings*, a view supported by McCormick (1997). There exists a relationship between science, mathematics, and engineering and many people see the latter as simply being applied science. Vincenti suggests that if this view is adopted then *studying the epistemology of science should automatically subsume the knowledge content of engineering* (Vincenti, 1993, p. 4). This, as he later states, is not the case as engineering knowledge should be viewed as an autonomous body of knowledge, identifiably different from the scientific knowledge with which it interacts. Engineers use knowledge to design, produce and operate artefacts, actions that in essence define the term engineering, whereas scientists use knowledge to generate more knowledge (Vincenti, 1993). This implies that engineering knowledge refers to that used by engineers and scientific knowledge refers to that generated by scientists, the latter being the producers of the knowledge that the former uses. However, engineers do in fact use knowledge to generate further knowledge and should also be considered as producers, although this is not seen as their primary function or concern. When defining engineering knowledge there is some inclusion of knowledge from science and mathematics, but it really develops from the nature of the engineering activity, such as the design process. I will show that the very nature of such a process makes the definition of the knowledge base difficult as it includes both tacit and personal knowledge gained from experience, from practice (Vincenti, 1993; Bucciarelli, 1994; McCormick, 1997).

Engineers are predominantly concerned with the solving of practical problems and to do this they need the necessary knowledge, which both serves and develops from the practical problem solving activity. The engineering design process is seen as usually starting with the identification of a problem with the limitations expressed as constraints, these becoming

more sharply delineated, and eventually leading to a focus on the desired functionality of the finished product (Court, 1998; Vincenti, 1993). The primary role of the engineer is based on applying knowledge and experience to technical problems in order to produce suitable solutions. Vincenti (1993) categorizes engineering knowledge, in the context of what he calls *normal design*, under the following headings:

- Fundamental design concepts.
- Criteria and specifications.
- Theoretical tools.
- Quantitative data.
- Practical considerations.
- Design instrumentalities.

(Vincenti, 1993, p. 208)

I will discuss briefly each of these in turn. *Fundamental design concepts* refer to how a device works, how it operates and the fundamental concepts which define it. The engineer must know the operational principle of their device (Polyanyi, 1962) in order to carry out normal design and this provides a key point of difference between engineering and science, as it originates outside the body of scientific knowledge and exists to serve some innately engineering purpose (Vincenti, 1993). Operational principles are clearly not enough to enable the engineer to design the device, knowledge of the technical criteria related to the device and its use must be available. These *criteria*, the *technical specification*, constitute an important element of what constitutes engineering knowledge. Again Vincenti (1993) sees this as a point of difference between engineering and science in that scientists do not aim to satisfy rigidly specified, practical goals or tasks in their quest for understanding whereas this is very much the nature of the engineers role. So an engineer needs device knowledge and the personal knowledge of engineers must be based on the devices they are dealing with and the social context within which they operate (Buccarelli, 1994). It follows then that device knowledge should not just be linked with abstract concepts but should be embedded within the community of practice where the devices are being used.

Engineers use a wide range of *theoretical tools* that include scientific and mathematical methods and theories, and what Vincenti (1993, p. 215) refers to as intellectual concepts. The mathematical knowledge is seen as being formulae or calculative methods used for quantitative analysis and design, theories that have no physical content in themselves but are used by engineers, and mathematically structured knowledge that is essentially physical and comes from prior scientific knowledge. The intellectual concepts are concerned with the language of thinking, not only for the calculations that must be carried out but also for the conceptualization process that forms a key part of an engineer's activity. As Vincenti states,

> these concepts range from being highly scientific to intensely practical, from specifically mathematical to explicitly physical. (Vincenti, 1993, p. 215)

They would include basic concepts from science, such as acceleration, work, electrical resistance and voltage, and those of an engineering nature such as mechanical efficiency and feedback in control engineering.

The *quantitative data* relates to knowledge of the physical properties engineers require and Vincenti (1993) divides this into two categories, descriptive and prescriptive knowledge. The former is concerned with knowledge of how things are, the properties of substances, physical constants and processes. Mathematical theories are also descriptive, they allow calculation of how things perform under given conditions or using certain assumptions. Prescriptive knowledge is concerned with knowledge of how things should be to attain the desired goal, the process specifications. This would include operational principles and engineering specifications that prescribe how a device should be for it to be able to meet its functional requirement. This differs from the design concept because it deals with the structure of the device itself as opposed to its performance, its function.

The *practical considerations* are concerned with knowledge gained from experience, from practice. This reflects the notion of personal knowledge and the emphasis on practice (Bucciarelli, 1994; McCormick, 1997).

Vincenti (1993) argues that this personal knowledge is learned on the job rather than from books or in the classroom. It also involves the use of judgement, which is more likely to be sound if it is based on knowledge gained from experience. Engineers learn from their mistakes, and success, just like everyone else. Vincenti (1993) believes that the use of this personal knowledge is essential to engineering and from it can be developed rules of thumb that allow rough checks of calculations or designs. This rule of thumb method is referred to as heuristics. Rules of thumb that have been developed from experience are used to formulate design solutions. For example the amount of heat required by a building in winter can be roughly calculated based on the volume of the building and a figure of 30 W/m^3 (CIBSE, 1999) thus reducing the difficult and protracted calculation of heat loads for a building to a simpler, solvable problem. A novice would be able to use this method but would not have the experience to judge how accurate the approximation was and would therefore lack confidence in using it, perhaps preferring to carry out the complete calculation.

Design instrumentalities are concerned with the procedures, the ways of thinking and the judgemental skills by which the process is carried out. Most engineering design work requires knowledge of the structured procedures that need to be followed in order to achieve the desired outcome. It also requires certain mental processes, ways of thinking, which may be derived from the intellectual concepts discussed earlier, by analogy, or by visualization. Vincenti (1993) suggests that efforts have been made in engineering education to teach both analogy and visual thinking in the classroom, however he sees practical experience as being indispensable and that knowledge of how to exercise judgemental skills, like knowledge to think visually, is mostly tacit. Vincenti (1993) concludes that such skills can only be learned through practical experience and sums up by stating that *knowing how to exercise such skills is, as much as anything else, what separates an outstanding engineer from an ordinary one* (Vincenti, 1993, p. 222). Vincenti (1993) is defining the difference between a novice and an expert in terms of how they view the problem. Glaser (1993) argues that in order to solve a problem the engineer represents the problem as a cognitive structure corresponding to it, based on domain-related knowledge and its

organization. By constructing this representation, the problem solver attempts to understand the nature of the problem and it is the quality, completeness and coherence of this representation that determines the efficiency and accuracy of further thinking. The difference between the expert's knowledge and novice's knowledge is that the novice's is organized around the literal objects given explicitly in the problem whereas the expert's is organized around principles and abstractions that subsume these objects. Experts derive these principles from their knowledge of the subject matter and they also have knowledge of the application of what they know, tacit knowledge, which together comprise tightly connected schema. Novices however lack this tacit knowledge and their difficulty in solving problems can be largely attributed to this rather than to the limitations in their processing capabilities such as the inability to use problem solving heuristics (Glaser, 1993).

I have discussed Vincenti's six categories individually but this does not imply that they should be taken to be distinct from one another. It is recognized that they *interact intimately* and that *though engineering knowledge has many threads, it is itself a tightly woven fabric* (Vincenti, 1993, p. 223). The knowledge used by engineers to solve problems must come from all the categories; they do not follow a hierarchical pattern, the engineer will move up and down within the categories and also back and forth between them. I have presented a view of the theoretical construct engineering competence that reflects Haertel's (1985) model, i.e. curriculum content, cognitive processes and behavioural outcomes. The curriculum content comes from the domains of mathematics, science and engineering and fits into the categories of the fundamental design concepts, criteria and specifications, theoretical tools and quantitative data. Both the cognitive processes and behavioural outcomes fit into the practical considerations and the design instrumentalities categories.

It can be seen from an engineering taxonomy, taken from the Engineering Professors' Conference (1989), that attitudes and values are particularly important for professional and vocational education. This taxonomy identifies not *levels* of learning but *types* of learning i.e. knowledge, skills,

understanding, know-how, and attitudes, values & personal qualities, and links them with teaching in the form of resources, process and assessment of outcomes. Imrie (1995) states that the two domains, cognitive and affective, are represented with the emphasis on *understanding* that *involves grasping concepts and being able to use them creatively* (*ibid*, p. 182). Both the Army and BTEC define the curriculum in terms of *objectives* and the taxonomy most widely used for the assessment of objectives is that developed for the cognitive domain (Imrie, 1995). Taxonomies have also been developed for the psychomotor and affective domains. The domains are defined as follows:

> The *cognitive domain* is concerned with knowledge and information and is subdivided into levels of cognitive ability (lower to higher): knowledge; comprehension; application; analysis; synthesis; evaluation.
>
> The *psychomotor domain* is concerned with the performance of skills: readiness; guided responses; mechanism; complex response; adaptation; origination.
>
> The *affective domain* deals with areas of learning that include such concepts as receiving; responding; valuing; characterization; organization; conceptualization.
>
> (Imrie, 1995, p. 177)

I do not intend to go any further into the process of domain specification for two reasons. The first reason is that it would present me with what I consider to be an enormous task, trying to specify all the domains clearly and unambiguously is not what I want to do in my study on validity. Such studies have been carried out (see Baker, 1992; Shavelson *et al*, 1992). The second reason is that I do not intend to conduct a quantitative study concerned with a statistical correlation between the definition and the assessment of the construct. I want to explore the *process* by which the construct is defined and assessed. In this chapter, I have discussed concepts of validity, defining and assessing competence in general terms and, more

specifically, defining the construct engineering competence. In terms of assessment of engineering competence and the validity of this assessment, the use of the assessors' judgement is, I believe, a fundamental issue. I will show in my study the lecturers are responsible for assessing student performance and the use of their judgement is linked to all of the issues discussed previously. It is the issues of assessors' judgement that I will now discuss.

Assessors' Judgement

It is widely accepted that competence-based assessment is a judgemental model (Gipps, 1995; Hager *et al*, 1994; Hager & Butler, 1996; Hambleton & Rogers, 1991; Wolf, 1995). Concern has been expressed over the use of assessors' judgement in terms of objectivity or bias when assessing student performance (Fleming, 1999; Gipps, 1995; Hager *et al*, 1994; Kane, 1994; Wolf, 1995). Assessors' judgement is likely to be based on their own interpretation of the performance criteria and of the evidence of student performance in relation to their interpretation (Fleming, 1999; Wolf, 1995). Despite the threat this poses to the validity of the assessment, attention has tended to focus on the issue of domain specification rather than that of assessors' judgement (Wolf, 1995). If the assumption is made that the assessment criteria are so clearly specified that performance can be judged explicitly against them then assessors' judgement is not such an important issue after all. However the difficulties with clarity and transparency when specifying the domain have already been discussed and this, coupled with the different contexts in which competence is assessed and displayed, means that assessors' judgement should be recognized as being equally as important as domain specification (Wolf, 1995). The use of decentralized assessors and devolved assessment methods can create a situation where assessors are advised or encouraged to collect any mix of evidence that enables them to make a valid judgement of competence. Such a situation does not guarantee common standards, allows for compensation and raises issues in terms of the validity, reliability and comparability of the assessment (Eraut, 1994, Gipps, 1995; Wolf, 1995). With regard to reliability and comparability, the role of validating bodies such as BTEC in

maintaining standards is considered crucial (Benett, 1993). BTEC (1988) views the maintenance of standards as being the collective concern of course teams that become aware of national standards through contact with the BTEC moderator.

Judgements about competence depend on the tacit knowledge and expertise of the assessors. They

> operate in terms of an internalized, holistic set of concepts about what an assessment 'ought' to show, and about how, and how far, they can take account of the context of the performance, make allowances, refer to other evidence about the candidate in deciding what they 'really meant', and so on. (Wolf, 1995, p. 67)

It is suggested that due to the wide margin of interpretation given to the assessors, who I will show in my study are the lecturers, that

> through their concrete evaluative decisions, they operationally define what is the valid knowledge of a discipline. (Kvale, 1996, p.220)

Kvale (1996) suggests that the assessors, the lecturers, thus decide what is true or false, right or wrong, and valuable to be developed further. If the assessors judge performance against their own standards rather than the laid down criteria then the validity, reliability and comparability of the assessment will be questionable. Training of assessors is seen as having a critical part to play in dealing with the problems posed by the use of assessors' judgement in assessing student performance (Burchell *et al*, 1999; Dunbar *et* al, 1991; Eraut, 1994; Linn, 1993; Kane, 1994).

Since judging student performance involves assessing multiple outcomes and numerous criteria, the results of the assessment must be collated in some way. The detailed performance profile of each student must be collapsed into a single reporting figure or grade (Gipps, 1995; Hambleton & Rogers, 1991; Wolf, 1995). This process is known as aggregation. The

simplest aggregation models operate by defining an overall percentage pass mark or a pass mark for each element of assessment where all elements must be passed. Either of these aggregation methods requires decisions to be made by the assessors and there is evidence that compensation by the assessors also occurs in the aggregation process (Smith and Shepard, 1988; Wolf and Silver, 1986; Wood *et al*, 1989). The degree to which the assessors use their judgement in the aggregation of results depends on their own experience, their familiarity within the field of concern and the degree to which they have developed their own internalized model of competence (Wolf, 1995). In criterion-referenced assessment the process of aggregation is considered to be best carried out by individuals who are familiar with the test content, understand the process by which the standards are set, have access to performance data, and understand the social and political context in which the tests are being used (Hambleton and Rogers, 1991). As competence-based assessment is a specialised form of criterion-referenced assessment then I consider the above applies equally as well to both.

The use of the assessors' judgement in defining the construct engineering competence, using internalized, holistic models to measure student performance, and aggregating the results of the assessment is an important part of my study. There are two reasons for this. The first reason is that I believe assessor's judgement should be given equal importance with the issue of domain specification. The second reason is I want to explore how the lecturers in my study use their judgement in assessing student performance and the impact this has on the validity of the assessment. The consequences of the use of the assessors' judgement will have implications for both the stakeholders at the RSME and for others working in competence-based education and training.

Summary of this Chapter

I began this chapter by exploring notions of validity and explained my rationale for adopting the unitary concept, as prescribed by Messick (1989), with construct as the unifying theme. As my study is concerned with exploring the nature of validity of the course assessment, I defined the

theoretical criteria against which the validity of the assessment will be evaluated. The criteria were focused on construct representation in the assessment and the claims made for the assessment.

In my study, the students that pass the assessment are considered to be competent to do the job of a Clerk of Works engineer. The assessment measures occupational performance from which inferences are drawn about the engineering competence of the graduates. I was concerned with competence-based assessment, a specialized form of criterion-referenced assessment. I have shown that defining competence is far from easy, the definition must contain the very essence of professional practice and I discussed the problems associated with domain specification and the assessment of the construct competence. What became evident was that a clear understanding of the construct, engineering competence, with performance criteria that have been derived from the key purposes and functions of the job was required. I decided against trying to specify the domains clearly and unambiguously in my study, as I wanted a broader definition of the construct. Using mainly Vincenti's model, I identified the basic categories that make up the construct engineering competence.

The model showed that the assessment of engineering competence must measure elements of the subject, i.e. mathematics, science etc., the cognitive and the behavioural domains. I discussed the issue of assessors' judgement and the use of their own internalized, holistic model of competence to define the construct, compensate for student performance and aggregate results. In my study I will look at the impact of the use of assessors' judgement on the validity, reliability and comparability of the assessment. I will explore the impact not only by gathering evidence from the lecturers themselves but also from other sources i.e. the students, serving Clerks of Works and their employing officers.

Before I present the data and its analysis I need to discuss and explain the rationale behind my choice of methodology and methods. The next chapter details the methodological issues and the methods used in gathering and analysing the data contained in my study.

CHAPTER 4

Methodology & Methods

Methodology

The Qualitative versus Quantitative Debate

Much has been written on the issue of whether educational research should be quantitative or qualitative in character. Quantitative research is concerned with *quantification* and *measurement*, and hence the collection and statistical analysis of numerical data. One of the criticisms levelled at quantitative research is that, although the numerical evidence produced appears authoritative, there are some fundamental doubts about its validity. Critics of this method also argue that it fails to take into account the very nature of human social life. The assumption being that it consists of *mechanical cause-and-effect relationships rather than contextually variable and complex processes of interpretation and negotiation that do not have determinate outcomes* (Bird & Hammersley, 1996, p. 15). The criticism is concerned therefore, not just with the numerical methods, but also involves disagreement over the nature of human behaviour and how it can be understood. Qualitative research emphasizes the importance of understanding how people's perspectives shape their actions and how diverse these perspectives may be. This is often seen as requiring an exploratory approach in which the researcher must suspend his or her own assumptions of the world in order to see the others' point of view. Actions are seen as being constructed over time and in ways that are sensitive to their context (Bird & Hammersley, 1996). Quantitative researchers have responded to the criticisms and, in the process, levelled their own criticisms against qualitative methods. The debate goes on, but little is to be gained in this study by discussing further the differences between what are fundamentally two opposed methods and philosophies of social and educational research. This is especially so as there is widespread agreement that both methods can be used to complement each other when carrying out

educational research (Bryman, 1992; Bird & Hammersley, 1996). There is a great deal of educational inquiry that uses both qualitative and quantitative methods (Bird & Hammersley, 1996) and although the initial intention was to conduct a qualitative case study, as my study evolved it became clear that it would involve the use of both methods. The discussion that follows focuses on the nature of qualitative research methods however where appropriate quantitative methods will be discussed together with the rationale for their inclusion.

Why a Focus on Qualitative Research?

In terms of educational inquiry, the sociology of education and curriculum evaluation research were two areas that turned to qualitative methods (Bird & Hammersley, 1996). Qualitative research can take many forms (Atkinson *et al*, 1993) and Bird and Hammersley (1996, p. 16) suggest it has, in general, the following characteristics:

- A strong emphasis on exploring the nature of particular educational phenomena, rather than setting out to test a pre-defined hypothesis.
- A tendency to work with 'unstructured data' i.e. data that have not been coded at the point of collection in terms of a closed set of analytical categories or a formally constructed scale. For example, when interviewing, open-ended questions will often be asked rather than questions requiring choice from pre-specified answers of the kind typical of postal questionnaires. In fact, qualitative interviews are often designed to be close in character to informal conversations.
- Generally, a small number of cases will be investigated in detail, rather than any attempt to cover a large number.
- The analysis of data involves explicit interpretation of the meanings and functions of human actions, and mainly takes the form of verbal descriptions and explanations. Quantification and statistical analysis play only a minor role, if used at all.

These characteristics can be reflected in the main features that Woods (1996, p. 82) sees as being present in most forms of qualitative research. These are:

- A focus on natural settings.
- An interest in meanings, perspectives and understandings.
- An emphasis on process.
- A concern with inductive analysis and grounded theory.

So how does this fit with my study that sets out to explore the nature of the validity of assessment at a military engineering training establishment? My desire was to explore how the construct engineering competence was defined, how this was assessed, and the nature of the validity of that assessment. As Bird and Hammersley (1996) suggest of qualitative research, it involves maintaining an open mind and fostering curiosity in order to delve beneath the surface to peel back the layers and reveal the complexity of the situation or process under scrutiny. Woods sees the concern with the process as showing an interest in *how understandings are formed, how meanings are negotiated, how roles are developed, how a curriculum works out, how a policy is formulated and implemented* (Woods, 1996, p. 87), and these concerns are reflected in my own study. He states that this has been described as *thick description* in which the *voices, feelings, actions, and meanings of interacting individuals are heard* (Denzin, 1989, p. 83). My study sought to discover the lecturers' perspectives when assessing the students, how they have interpreted the way in which the course assessment has been devised, the knowledge required by the students, and how this was demonstrated. Other perspectives were also sought, i.e. students, the BTEC moderator. Throughout my study the aim was to focus on exploring the assessment process in order to discover the nature of its validity. As my study progressed, it became clear that quantification, the use of numerical tables and simple numerical comparison would have a role to play in analysing some of the data. There was a move away from the interpretive nature of qualitative research and the final section of this chapter, analysing the data, will deal with this issue in more detail.

Why a 'case study'?

Yin (1994) argues that the most appropriate strategy for research questions involving the how and the why, as with my research, is the case study. My study explores not just how competence is defined but also asks why it is defined in this way. It is a study into the views held and judgements made by staff and students on what constitutes military engineering competence and how it is assessed. It also asks in what ways assessment of this construct is valid. My study does not set out to examine student assessment results but to explore the process of assessment, what is looked for, why this is important and how this is being done. It sets out to explore what is happening and to offer reasons as to why this is so: how is the construct defined and, in relation to this, how valid the assessment actually is, and if possible, to explain the reasons why. Therefore I believe the main purpose of my study to be to *explore or investigate little understood phenomena or behaviours and discover the important underlying patterns, themes, and factors which affect them* (Falkner *et al*, 1993, p. 17). I consider this necessary because the issues explored in my study are I believe, rarely investigated and little understood, a view supported by the lack of similar studies with which to compare my own.

In order to address these issues I decided that I should carry out the research at the Royal School of Military Engineering (RSME). This lead to the selection of the case study as the appropriate form for the research. The decision to undertake the research at one institution, the RSME, does not in itself define the case to be studied. What is needed now is to define the unit of analysis the definition of which is generally seen as being related to the way in which the initial research questions have been defined (Yin, 1994). The unit of analysis can be defined in one of two ways, as the establishment or as the engineering courses. As this is a military establishment, the course lecturing staff are a mix of military and civilian personnel, there is a strict hierarchy, a code of discipline, and all the students are Royal Engineers who have undergone an intensive selection procedure. Thus the culture of the establishment is clearly different to that of a civilian college. However its aim is arguably the same as the civilian colleges, i.e. to produce HND level

engineers who can perform competently in the workplace. The courses are designed to produce a military engineer. They are designed and administered by the military. The teaching and learning environment and the process of assessment are thus embedded within the culture of the establishment. This is the fundamental difference between the courses in my study and those carried out at a civilian establishment. The fact that the courses lead to the civilian qualification, the BTEC HND, means that the assessment process should parallel, or be equivalent to, that of civilian colleges. The range of assessment methods used i.e. written examinations, projects, presentations, laboratory and homework assignments, is not influenced culturally however this cannot be said of the style, content and process of assessment. The specific purpose of the assessment of the students is to measure their ability to perform as military engineers, their competence, and this is what it is designed to do. Whereas, within a civilian establishment the purpose is likely to be generalized, i.e. it would not relate to a specific working, or cultural, group.

My study explores issues of the assessment of engineering competence and its validity within a military culture. In the previous chapter I discussed the theoretical development of the construct, engineering competence, how it should be assessed, and that construct validity should be sought. In order to look at these issues in practice, the focus must be on a military engineering course or courses that enables access to the assessment used, the lecturers who design, administer and mark them, the students, and the course sponsors, i.e. BTEC and the Army. The unit of analysis is thus defined as the BTEC HND Clerks of Works courses in building services engineering that are managed by the Electrical and Mechanical Branch at the RSME. The courses are designed to satisfy two criteria, to award students the BTEC HND and to certify them competent to carry out the duties of a Military Clerk of Works. There are two streams that will be researched, the electrical and the mechanical. They were chosen because they parallel each other in all respects, with the only difference being the subject content of the engineering modules, i.e. electrical or mechanical.

Criteria for Assessment of Research - Reliability, Validity and Relevance

There are three criteria against which my research should be critically evaluated. These are reliability, validity and relevance (Faulkner *et al*, 1993; Bird & Hammersley, 1996). In order to give credibility to my research it is necessary first to explore what is meant by these criteria and then to show how they were dealt with in my study.

A study would be considered to be *reliable* if it could be argued that if it were repeated the results could confidently be predicted to be the same. For reliability to be present the research findings should be similar if the research was repeated on the same people, or different people from the same group, at a different time either by myself or by others. In my study, due to the nature of the military environment where military personnel change every two years or so, different lecturers holding the same post were interviewed. Different student groups were surveyed and also interviewed, and therefore it could be argued that some reliability exists already as the responses were indeed similar. The issue of reliability of the methods of data collection, i.e. the questionnaires and interviews, is also important. The student questionnaires were piloted and the responses were reviewed in order to reduce ambiguity in the questions and hence to make the answers more reliable and, indeed, valid. Compiling and using the questionnaires proved to be a difficult process and this is discussed later, however I would expect the data collected to be very similar if future student groups were surveyed using these questionnaires.

With regard to the interviews the issue is not quite so clear. As my study explored the lecturers' views and practice, it was evident that the interview process generated new thoughts and ideas about the methods of assessment used, how they were marked, and what was actually being assessed. Issues such as what performance criteria were being assessed and what were the most appropriate methods to do this were discussed at length with the lecturers. It became evident when the lecturers were interviewed for a second or third time their views, and in some cases practice, were changing

because they had reflected on the discussion from a previous interview. If this is so, would the data collected be similar if collected in, say, two years time? Due to the nature of the issues being explored and the fact that all four military lecturers changed post at least once during the course of my study I believe the answer to this question would be a definite affirmative.

Validity in this context implies *truth* and there are many underlying questions that this notion of truth brings with it (Bird & Hammersley, 1996). It is not the intention here to enter into a philosophical debate about the nature of truth but to discuss in what ways my study can be recognized as being valid. There are three main features upon which the validity of qualitative research usually rests: unobtrusive measures; respondent validation; and triangulation. As a senior lecturer at the RSME I already knew a considerable amount about its methods and practices and I made every effort to maintain an analytical distance thus rendering the situation anthropologically strange (Open University, 1994). However, although I tried to limit the influence my research was having on the existing situation, this proved to be difficult. There were a number of instances where, as a result of interviewing lecturing staff and allowing respondents to read the transcripts and the results of the initial study, processes and practices were subsequently reviewed and modified. However, feedback from the respondents showed that they felt that a true picture was being presented in my study. In terms of triangulation, I believe this is evident from both the means of collecting the data, i.e. interviews, questionnaires, observations, and documentary analysis, and from the sources of the data i.e. the lecturers, the students, the BTEC moderator and the documents.

Finally, to be of value the research findings must be *relevant* as well as valid. The results of the study must be of significance to the intended audience. In my study this audience consists of a number of different groups. The first is the participants of the study and in particular the lecturing staff. I have no doubt that this study was of extreme interest and relevance to them. It offered them a review of the structure, content and assessment strategy for their courses and encouraged them to discuss and reflect on their own practice, they became reflective practitioners. Next

there is the Army and BTEC. They are concerned with two issues, that the course content meets their requirements and that the assessment strategy ensures the courses produce competent military engineers, at HND standard. My study will also be of relevance to a wider audience of educational practitioners and specifically those involved in the training of professions such as engineering, where assessing competence is critical. The reason for this is that I believe that my research contributes to the understanding of the assessment of competence and of how issues of validity of assessment can be identified and explored, particularly in terms of the unitary concept.

The Issue of Generalisation

Although validity of engineering assessment should be an issue of concern for all institutions of further and higher education, it would be an impossible task for the research to encompass all such institutions. Feasibly, a small number of institutions could have been chosen, the Royal Electrical & Mechanical Engineers, the Royal Air Force and the Royal Navy all conduct similar training, but the nature of my research lends itself to investigation within a single institution. It does not, at this point, involve the relationships between either similar military or civilian establishments. However, the expectation is that the assessment used by these institutions, and the types of validity they would expect to be present, are mirrored by those of a single institution such as the Royal School of Military Engineering. This raises the issue of generalisation, which is seen as a problem with qualitative case study work (Schofield, 1993).

Schofield (1993) argues that the view of generalisation, shared by many qualitative researchers [he quotes Guba & Lincoln (1981, 1982); Geotz & Le Compte (1984); and Stake (1978)] involves a number of areas of consensus. One area is that generalisability, producing laws that apply universally, is not a useful goal or standard for qualitative research. However, the idea that studies in one situation can be used to speak to, or help to form a judgement about, other situations is not totally rejected by qualitative researchers. This depends on the descriptions of the site being studied and those situations to which generalisation is to be made. These are

crucial in allowing the search for similarities and differences between the situations. Analysis of these similarities and differences thus makes it possible to make reasoned judgements about the extent to which the findings from one study can be used as a working hypothesis about what might occur in another situation. This can be best thought of as the fit between the study and other situations in which the concepts and conclusions could be applied. Schofield also argues that studying a situation with special characteristics, such as in my study, does not necessarily restrict the application of the findings to other similar situations. This depends on how the findings are linked to the special characteristics, how deeply are they embedded.

I will show that my study does cover issues and develop concepts that can be applied to other situations. There are similarities that can be identified between the RSME and other institutions and it is my belief that these result in a good fit. This belief comes from my experience in secondary education and from working as an engineering officer in the Royal Air Force, in industry as an engineering manager, and with the Royal Engineers and the Royal Electrical & Mechanical Engineers as an educator. By exploring the issue of competence, assessment and validity at the RSME I will show that this does supply valuable information about these concepts that could be used in the theory and practice of education generally.

Methods

The Ethical and Ethnological Issues

The ethical issues were really concerned with how I would deal with the effect of my study on the people involved. I set out to make it clear to all the lecturers that my intention was not to pass judgement on their processes or practice but was to explore what they were actually doing, how the process of assessment was carried out. I also made it clear to all the participants that they would not be personally identified in any publication of my work and every attempt would be made to ensure their anonymity. However, in the case of the Senior Military Lecturers I pointed out that it was likely that

with a little effort they might be identifiable. As part of the interview process I asked for the interviewees' permission to tape the interviews, which was agreed in all cases. I made sure that everyone was given access to the taped recording of their interviews and access to any written material I produced for publication. I discussed any inclusion with those concerned, especially the Senior Military Lecturers. I also gave all participants the option to withdraw from my study if they felt, for any reason, this was necessary. The same principles applied to the students and the BTEC moderator. None of the participants felt there were any problems with these issues and hence none felt it necessary to withdraw from my study. To ensure the points were clearly expressed, everyone that participated was given an ethical statement to read, and keep as a record, prior to their participation. A copy of the statement for staff and students is included as Appendix A.

The ethnological issues were important because my study took place at a military establishment. The first issue concerned obtaining permission to carry out the study as it was likely that some of the material I would be dealing with would have some restriction, such as management in confidence, placed upon it. I wrote to the Commanding Officer at the beginning of my study outlining its nature and purpose and seeking permission to undertake the research. My request was supported and permission was duly granted. The second issue was concerned with the distinct hierarchical nature of military establishments and the degree of rank consciousness that permeates such a culture. As a civilian Senior Lecturer I held an equivalent officer rank and all those I interviewed, both military and civilian, were therefore subordinate to me in rank but I was not part of their line management chain. As with the ethical issues above, I was concerned that everyone understood the nature and purpose of my research and did not feel in any way that they were being subjected to performance assessment. I needed to ensure that the interviews were not seen as interrogation but that they were open and relaxed such that the interviewees could freely express their views. I feel confident that this was achieved.

I tried very hard to ensure I remained anthropologically strange, as mentioned earlier, and throughout the process of conducting the research I made every effort to ensure I did not bias any of the information presented to me. It was clear that my study did impact on the processes and practice employed by the lecturers but I would argue that this occurred as an indirect result of the study and not from any direct, intended influence on my part. I was carrying out a process of which, in my opinion, reflective practice was an unavoidable consequence and what I tried to do was to note any changes as my study progressed.

Collecting the Data

I will now focus on how I collected the data. The data were collected in four ways, from: documentary analysis; interviews; questionnaires; and observation. I identified five groups of participants that were concerned with the teaching, learning and assessment of the students. These were the mechanical engineering lecturers, the electrical engineering lecturers, the science and computer lecturers, the students and the BTEC moderator.

The *documents* used in my study came from BTEC, the Army and internally from the RSME. As the courses were designed to train military engineers and also carry the award of the BTEC HND they should incorporate both the Army training requirements and those of BTEC. Thus data came from documentary sources in the form of Army Training Organisation pamphlets and BTEC guidelines. The former was produced by the Army Training Organisation to be used throughout the Army and give the general guidelines on training. They were designed to cover the setting of course training objectives, course design, testing and recording of results. The BTEC publications cover all aspects of course design and assessment. Through the process of documentary analysis described later, I will show how they each defined the concept of validity and how they expected it to be incorporated in the assessment process.

The assessment of the foundation phase, the electrical, and the mechanical subject modules was comprised of homework assignments, laboratories,

presentations, examinations, and design projects. It is from these assessment documents and the associated interviews with the lecturers that the key question of what is actually being assessed will be answered. This will also result in an exploration of the use of the lecturers' expert knowledge in making judgements on what is assessed. I will also show what knowledge, skills, and abilities were being assessed formally or informally within the process. The data from the students showing what they believed was being assessed will add substance and support to this argument. A secondary data source used in the study was the *Review of Royal Engineers' Clerks of Works Employment and Training* (TDT, 1999). This document contained the results of a survey of over 80% of all Clerks of Works (240 excluding current students), most of their Commanding Officers and of interviews with a cross section of both populations. This document provides a source of data on how the courses are perceived to have prepared the Clerk of Works engineers for their role in the Field Army.

Interviews were carried out with participants and were semi-structured. These are the most favoured by educational researchers as this type allows respondents to express themselves at some length but offers enough shape to prevent aimless rambling (Wragg, 1994). This, and the intended open-ended nature of the questions, allowed for the exploration of themes during the interview whilst still maintaining a basic structure. A short extract from an interview with one of the military lecturers (ML8) is shown at Appendix B. In collecting data via the interview process I was exploring the issues and seeking explanations for what I was discovering as the study progressed. I was carrying out a process of progressive focusing. For example the interview at Appendix B was followed by a second interview with this lecturer that developed some of the issues discussed in the first.

The mechanical and the electrical engineering lecturing staff consisted of a Senior Military Lecturer, a military Lecturer, a civilian Senior Lecturer and a civilian Lecturer for both subject areas. The civilian Senior Lecturer and lecturing staff from other departments input into the teaching of the courses but they played relatively minor roles in the delivery and assessment of the courses. The Senior Military Lecturer was the Course Director and together

with the military and civilian lecturers, designed and administered the courses. These were the key staff in that they set the course curriculum, teach and assess all the modules in the engineering phase of the courses. In theory this meant that three electrical and three mechanical lecturers would be interviewed at least once. However because of the posting of some of the military personnel during the duration of my study, there were in fact nine staff interviewed, five of which were interviewed twice. The interviews of these staff were designated L1 - L14. In Chapter 5 I will make it clear when I quote from the interviews whether the lecturer was military (ML) or civilian (CL).

The science and computer lecturers were responsible for teaching the foundation modules for both the electrical and the mechanical courses. This covered the mathematics, science and basic computing skills which were taught in the first few months of each course with the aim of bringing all the students to the required level of knowledge and understanding in these subject areas. The knowledge base of this phase consisted of the traditional subject matter of mathematics and science. I felt it was important to get this data as the view that the students require a good level of understanding in the principles of mathematics and science was evident from my review of the literature. Two civilian lecturers and one military lecturer were interviewed from the Science & Computer Department, two were interviewed twice and their interviews were designated CL15-CL17, ML18 and ML19. They were all involved in the delivery of the mathematics and science foundation modules, and the higher mathematics module to both the electrical and the mechanical courses.

In order to gain the views of the *students* I decided to issue two questionnaires to them, one at the end of the foundation phase and one at the end of the course. Since each course runs for two years, this resulted in a running programme from January 1997 until February 2000. The *Initial Questionnaire 1* was designed to collect data on the students perceptions, the purpose and appropriateness of the assessment and whether it measured their potential to perform well as engineers, after they had completed the foundation phase of their course and were a few months into the engineering

phase. The *End of Course Questionnaire 2* was designed to re-visit some of the earlier questions but with the emphasis on the final project. This allowed for the student perceptions to be re-assessed at this point, prior to them going out on attachments to civilian industry. The design of the questionnaires presented some problems highlighted by the realisation that *the effective use of questionnaires does indeed demand a clear understanding of the overall research context* (Youngman, 1994, p. 248). A copy of each questionnaire is included at Appendix C and D. In order to explore the responses from the questionnaires I carried out four student interviews following the return of the end of course questionnaires. These were semi-structured interviews, designated S1 - S4 and a general interview schedule is included at Appendix E. I decided that the questionnaires would capture the views of the students at two points on the course and this, coupled with a small number of confirmatory interviews, would give me sufficient data in this area. I felt that the task of interviewing all students twice would have proved too immense and time was an issue for both myself and for the students.

Of the two interviews with the *BTEC Moderator* one occurred very early on in the initial study and the other towards the end of my study. They were designated as interviews B1 and B2. The visits of the moderator were governed by the need to have the students' work moderated and therefore the timing of these interviews was outside of my control. As a result of these visits he submits a report to BTEC and to the Army, via the Senior Military Lecturer, containing his findings and I obtained copies of the reports from May 1996 up to March 2000. This data supported that collected from the interviews with the BTEC moderator and was collected to explore the curriculum and assessment issue from the BTEC perspective.

As well as the interviews, a number of *observations* took place. The first of these was a meeting between the mechanical and electrical staff to discuss the way each course was taught and assessed. The purpose of this meeting was to discuss any differences in approach to assessment and grading of students and thus to develop commonality between the two subject areas. This meeting occurred approximately mid-way through my study and it was

admitted to me that it was set up as a direct result of my discussions with the staff. The second observation was the student presentations to the mechanical lecturers as part of the final project. The aim here was to observe how the staff discussed and marked these presentations, what was being measured and how this was being assessed e.g. were any personal qualities being assessed and if so what and how?

During the earlier interviews it was stated that meetings of the assessment teams concerned with the grading of the final projects, i.e. the three key staff, took place towards the end of the courses. It was suggested that both the electrical and the mechanical departments held these meetings. Having attempted to observe these meetings it transpired that in reality they did not occur. The new electrical Senior Military Lecturer decided he alone had responsibility for grading the final projects and the mechanical lecturers discussed it informally amongst themselves. This was unfortunate as I felt that had I been able to observe the meetings implicit criteria may have been voiced that would not necessarily come out during interviews. In an attempt to replace the data I would have obtained, I set up an interview (ML20) with the two mechanical military lecturers to discuss how they had arrived at the final marks they awarded for the final projects. This was a retrospective reconstruction of some of the discussion that *might* have taken place.

Analysing the Data

As my study progressed so the amount of data increased exponentially and this presented me with the need to identify the best way to analyse it. I first considered the two ways I could have analysed my data to be by deductive content analysis or grounded theory arising from inductive content analysis (Easterby-Smith *et al*, 1994). The former is a method that involves analysis by number and frequency whereas in the latter the researcher goes more by feel, by intuition, producing themes and patterns from the data that can be used as the basis for interpretation. Predictive theories and a deductive approach may then be developed from this form of analysis (Bird & Hammersley, 1996). The differences between these two methods are represented clearly in Table 4.1.

CONTENT ANALYSIS	GROUNDED THEORY
Bitty	Holistic
Go by frequency	Go by feel
Objective	Closer to the data, open much longer
Deductive	Inductive
Testing Hypotheses	Testing out themes, developing patterns

TABLE 4.1 Differences between *content analysis* and *grounded theory*.

(Easterby-Smith *et al*, 1994, p. 345)

Having looked at these two methods, the use of either method on its own would not suffice. I had identified a number of research questions and thus established what I believed were the themes and the hypothesis I wanted to explore so this made the use of grounded theory less appropriate. However, some of the data generated for example from the interviews and observations did lend itself to the inductive approach. I considered the implications of using an intuitive approach to data analysis in terms of the nature of the organisation and people my study was directly relevant to, i.e. the Army. There is evidence that qualitative studies are often seen as being *impressionistic, subjective, biased, idiosyncratic and lacking in precision* (Bird and Hammersley, 1996, p. 101). Quantifying qualitative data, turning it into numbers, is a way to combat this perception as *numbers have a seductive air and sometimes, thinking politically of the acceptability of their findings, they* [researchers] *gear their data to quantitative statements* (Easterby-Smith *et al*, 1994, p. 344). These were issues that I had to be aware of and within the military culture there is no doubting that the use of numbers that would be more appealing to senior military officers and other military personnel involved in the study. Although I felt I still needed to maintain a *feel* for the data and to present a holistic view, a more deductive approach began to evolve as my study progressed. Contributing factors to the development of this approach were the military culture within which my study was embedded, the strictures of the Army Training Organisation and BTEC, and the nature of some of the data generated.

I identified my three sources of data as being documents, interview transcripts and questionnaires. In terms of the *documents*, there were two distinct types, those that were suited to interpretive analysis and those that were suited to simple numerical analysis. For the first type, e.g. the Army Training Organisation, the BTEC and the TDT documents, I decided to follow what Faulkner *et al* (1993) describe as the most common method, based on grounded theory. This was to read the documents and note down points of interest, teasing out the themes, patterns and categories but with the process being guided by the research questions. There are two distinct strands to this form of data analysis and these relate to the issue of the content of the document and any comment or interpretations that occur to the researcher about the content, or possibly the lack of it. The second type were those that were part of the actual assessment process, e.g. course assessment programmes and assessment methods. These documents were analysed using simple numerical techniques, frequency counts and percentages, because this was seen as the appropriate method in relation to exploring how the content of these documents helped to define the construct and establish the predictive nature of the assessment.

In terms of data from the *interviews*, I decided to identify basic categories starting with who the interviews were with e.g. science & computer lecturers, electrical lecturers, mechanical lecturers, students, the BTEC moderator. Then I identified sub-categories from my research questions e.g. validity, BTEC specifications, Army training objectives, the assessment process, judgements being made, views of knowledge, skills, attitudes and abilities. This enabled me to check the validity of my categories by comparing data from the different sources i.e. through triangulation. The data generated from the interviews was analysed using an interpretive approach.

The *student questionnaires* generated data in response to the questions that were divided into two categories, numerical responses and additional comments. The data gave an indication of the students' views on the course assessment process i.e. their perception of what they thought was assessed, how this was being done, and how well it predicted future performance.

Although there was an insufficient number of students, due to the size and duration of the courses, to carry out any sophisticated statistical analysis on student responses I felt there were sufficient numbers to look at frequency of responses and to use simple percentages. I also used an interpretive method similar to that of Miles and Huberman (1984) adopting the idea of a matrix developed from the questions used in the student questionnaires and linked to themes emerging from data generated from the interviews with the lecturers. The questionnaires were focused on gathering the students' views in the categories that came from the research questions and this analytical process identified the common views that were qualified by reference to the student interviews. The data, analysed in this way, gave me other evidence in relation to the lecturers' views on the assessment process.

Throughout the process of data analysis I was conscious of the need to follow a systematic approach to provide rigour for the academic and military assessment of my study without losing sight of the intuitive nature of qualitative research and analysis, the feel for what I was doing. My intention was to make the whole process as valid and reliable as possible and to find the right balance between rigour and intuition, but as has been discussed, as the study progressed intuition and interpretation began to give way to a more deductive approach in analysing the data.

CHAPTER 5

Presentation and Analysis of the Data

Introduction

My study was a case study carried out to evaluate the validity of the assessment of military engineering competence that focuses on the Clerk of Works courses delivered at the RSME. The process this chapter follows begins by discussing how the Army, in the form of the Army Training Organisation, BTEC and the lecturers deal with validity as a concept. In a similar manner, the next section deals with how the construct, military engineering competence, is defined in terms of the Army training objectives, the BTEC units, and the role of the lecturers in dealing with the Army and BTEC methodologies. What will emerge from these two sections will be the concept of validity explicitly designed into the assessment and a definition of the domains that are considered to define the construct. The next part of this chapter will focus on the assessment process i.e. the methods used and the use of the lecturers' judgement in assessing the students. The evidence collected will be used to measure the validity of the assessment against the theoretical criteria I described on page 26. These criteria were based on the unitary concept of validity and should deal with whatever type of validity is considered by the lecturers to be designed explicitly into the assessment. I will need to present other evidence about the assessment in order to ensure triangulation occurs. The other evidence will come from the students, serving Clerk of Works engineers and their employing officers. Throughout the presentation and analysis of the data reference, where appropriate, will be made to my research questions and the discussion of the literature presented in Chapter 3.

Conceptions of Validity

My first set of research questions were concerned with establishing the source of the validity of the assessment and the importance given by the

stakeholders of the need to state explicitly the form of validity required. My first task is to establish how the stakeholders in my study defined the validity of assessment and my second task is to explore what explanation or guidance was given on how to ensure this validity was present. In order to establish the source of the validity of the course assessment I need to look at how the lecturers referred to the BTEC and Army literature on designing valid assessment. I will then be able to state the source of the validity of the assessment and what form of validity is expected to be present in the assessment. I will begin with an evaluation of the BTEC publications followed by an evaluation of the Army Training Organisation publications.

BTEC (1986a) referred to validity, reliability and utility as characteristics of good assessment and it was stated that appropriate attention must be paid to reliability and utility, but validity had overriding importance. Validity of assessment was defined by BTEC as

> the extent to which it [*the assessment*] serves its purpose. For example, in assessing problem-solving skills, an item that requires students to solve problems will be valid, whereas an item that requires only recall of information will not. A valid assessment will therefore employ methods - and accord them relative importance - that reflect the aims and objectives of courses and units. (BTEC, 1986a, p. 2).

This was the only explanation of validity contained in the publication. It implied that the purpose of the assessment and the way in which it met the purpose must be appropriate to the aims and objectives of the course. The problem-solving example quoted indicated that BTEC were aware of the need for the assessment to measure the relevant performance criteria in order to be valid and that simple recall of information to solve a problem did not define problem-solving sufficiently as a domain. Although not stated explicitly BTEC appear to be recognizing the problem of domain specification. There is also the implication that different methods of

assessment must be used and that the choice of methods is dependant on the aims and objectives of the courses.

The BTEC moderator (B1) considered the assessments to be valid if the link between the course content and the course assessment was present and well established. It was confirmed that he considered that the course content reflected the aims and objectives for the courses. He felt that course teams had been set up to match course content to the course assessment and to decide on the type of assessment that should be used and it was the responsibility of the team to ensure the link between content and assessment existed. As far as he was concerned the lecturers were the experts and he considered their judgement to be sound (B2). In terms of methods, he listed the types of assessment contained in the BTEC assessment and grading publication (1986a): case studies; practical exercises; oral, aural and visual processes and presentation; role-play; and long and short answer questions. He stated that his initial concern over the predominant use of pen and paper examinations had been overcome through the use of more of the above methods and he was now satisfied with the variety of assessment methods used during the courses. He was concerned with the look of the assessment, that it looked right to experts in the field and therefore had face validity (B2).

Short definitions of reliability and utility were also given by BTEC (1986a). The definition of reliability matched the general view found in the literature. In terms of the relationship between validity, reliability, and utility, it was stated that *appropriate attention must be paid to reliability and utility, but validity has overriding importance* (*ibid*, p. 2). The only other BTEC guidelines given were concerned with how to optimize validity and reliability. To achieve this, they stated that sampling of student performance *should recur over a period of time, in a variety of contexts, and through a range of methods* (*ibid*, p. 2). The range of assessment methods was the same as those listed above. It was later stated that the extent of the sampling should be adequate, not excessive, and that the teaching and learning strategies were reinforced, not distorted, by the assessment strategy (BTEC, 1986a). This is hinting at the possible problem of teaching to the test.

However, it was unclear whether adequate sampling was seen by BTEC as dealing with the problems associated with domain specification, reliability, or both. The BTEC moderator (B1) stated he was concerned with the issue of reliability. What was clear was that reinforcement of integration, bringing together knowledge and skills from different parts of the course (BTEC, 1986c), was stated as a requirement to be addressed in three ways:

1) Across objectives in units, particularly in end-of-unit assessments;
2) Between units in a course, e.g. through assignments that bridge units;
3) Between the course and the job, e.g. through work-based assignments. (BTEC, 1986a, p. 5)

Although the BTEC publications did not state directly that project-based assessment would be the best fit they did see carefully designed projects as assessing across both objectives and units, and that they could be work-based. BTEC ensured the use of project-based assessment on the courses by their requirement for the completion of the Integrated Project Unit.

All the relevant BTEC publications were held in both the electrical and mechanical lecturers' offices. Although all the lecturers interviewed said they were aware of the publications, the only two that were familiar with some of the content were the two lecturers responsible for dealing with the BTEC requirements for accreditation of the courses in terms of the award of the HND. None were able to discuss the content of the Assessment and Grading General Guidelines (BTEC, 1986a) and therefore there was no evidence that their views of validity were influenced by the guidance given by BTEC.

The *Army Training Organisation* (1995) gave definitions of what they called fundamental terms and concepts that included validity. They defined a valid assessment as *one which measures what it is supposed to measure* (Army Training Organisation, 1995, p. 3). The implication was that the assessment should be designed specifically to measure the training

objective, and hence the construct, but this was not stated explicitly. Ideally, the assessment should be designed by the Training Design Team (TDT) as an independent body, guided and advised by the lecturers, however in particular cases where specialist knowledge was required the assessment could be designed and administered by the lecturers with only minor input from TDT. The Training Design Team at the RSME considered the Clerk of Works courses required specialist knowledge and the lecturers were left to use their judgement to design and administer the assessment.

The Army Training Organisation (1995) stated that validity had three components: *reliability*; *relevance*; and *range. Reliability* was defined in similar terms to BTEC but the definition ended with the statement that it was *possible for a test to be reliable without being valid, but a valid test cannot be unreliable* (*ibid*, p. 3). Although not the best phraseology, this statement was the justification for including reliability as a component of validity. I agree with the BTEC view that reliability is an individual concept that has a relationship with validity and this is supported by reference to reliability *and* validity in the assessment literature discussed on page 14.

Relevance was defined by a valid assessment measuring only *learning which is pertinent to the purpose of the test* (*ibid*, p. 4). It was stated that the test must not direct the students to carry out irrelevant performances nor should this be allowed to affect the scoring and the example given was the awarding of marks for writing style when only factual knowledge was being assessed. This did indicate that there was recognition of the threat from construct irrelevant variance and that effort should be made to deal with it. The term face validity was introduced here and was linked to the carrying out of irrelevant performances and defined as an assessment *appearing to measure what it is supposed to measure* (*ibid*, p. 4).

The third component of the Army Training Organisation definition was covered by the suggested requirement for assessing the full *range* of expected learning. It was recognized within the document that, in practice, it would rarely be possible to assess each element. To overcome this, it was suggested that some form of random sampling take place. Sampling was

seen to be the measuring of representative parts of the relevant behaviour and it would be a random sample if each element of learning stood an equal chance of appearing in the assessment. At first glance it appeared that this component was an attempt to overcome construct under-representation but for this to be so, great care would have to be taken when considering the randomness of the sample as it must include the important dimensions of the construct. The difference between this random sampling and the sampling discussed by BTEC was that the Army publications did not suggest explicitly the use of different methods of assessment.

All the lecturers interviewed were aware of the training and publications available but only three had attended a training course and none were conversant with the Army publications on assessment and validity. The only compulsory training for any lecturer was attendance at a Basic Lecturers course that focused on presentational skills. Unless lecturers attended the *Testing and Testing Techniques* course, or studied the associated pamphlet, they would not be familiar with the Army guidelines. As none of the lecturers had attended the course or had the pamphlet and could not discuss the concepts at interview, the evidence shows that they were not familiar with the guidelines.

Guidance was available from both BTEC and the Army Training Organisation on the validity of assessment. They both attempted to define validity and as a result they gave standard statements but not any explicit guidance on the need for construct validity. In fact the only reference to any specific type of validity was the Army Training Organisation's reference to face validity. They both indicated that sampling of student performance should occur, implying an awareness of the need to assess a representative sample of the construct i.e. to avoid construct under-representation. However in view of the discussion on conceptions of validity on pages 15 – 27, the guidance given by BTEC and the Army Training Organisation was limited. This presents two possible inferences. The first is that both BTEC and the Army Training Organisation considered the curriculum was designed, administered, and assessed in such a way as to lead to construct validity naturally being present. This would stem from a belief that the

lecturers based their lectures and assessment around the construct defined from either the job analysis via the training objectives, the BTEC unit specifications or a combination of both. Neither the BTEC nor the Army guidelines were written specifically for these courses and it would appear implicit in BTEC and Army thinking that validity will be present by virtue of the nature of the judgement of those designing and administering the curriculum and its assessment. This appears acceptable if based on the assumption that BTEC and the Army Training Organisation view the system within which the assessors operate as ensuring any judgements made would fall in line with the prescribed policy. The comments from the BTEC moderator discussed earlier show he supported this assumption. The second alternative inference is that the authors of BTEC and the Army Training Organisation publications on assessment had a limited understanding of the theory of the validity of assessment. This may well be the case based on the earlier discussion of available texts on the assessment practice on page 18 that showed issues of validity were dealt with somewhat summarily. However I decided that researching the second inference would not add anything of further relevance to my study.

Analysis of the data from the interviews revealed that none of the lecturers referred to or had detailed knowledge of either the BTEC or the Army guidance in relation to the validity of assessment. The lecturers did not refer to any of the types of validity described in Chapter 3. The lecturers' view was that the assessment of the course was valid because the graduates of the course could do the job of a Clerk of Works engineer.

> "The reputation of the end product [*the Clerk of Works engineer*] is extremely high." (ML2)

> "Experience has proven that a student that completes the course satisfactorily will generally prove to be a very good Clerk of Works." (CL10)

> "I think that anybody that gets through [*the course*] is more than capable of doing what the Army wants." (ML8)

There was no evidence that the process of assessment design had been conceived in a way that was likely systematically to encourage the development of valid assessment instruments. BTEC, the Army Training Organisation and the lecturers cannot be defined, therefore, as an explicit source of the validity of the assessment. There is no evidence of an explicit intention by any of the three to ensure the validity of the assessment. However I can still evaluate the validity of the assessment to encompass the Army Training Organisation and BTEC definitions of validity and the claims made by the lecturers about the assessment. To do this, I will need to explore the nature of the assessment process inherent in the culture of the organisation and the role of the lecturers as an implicit source of validity of the assessment. I recognize that the lecturers' claim is about occupational performance and the circumstances of my study present an ideal arena in which to conduct a longitudinal study of occupational success, such as the Royal Naval study conducted by Gardner and Williams (1973). This was beyond the scope of my own study. My study must progress by exploring the implicit nature of the validity of the assessment. This exploration will involve analysing the assessment methods and the application of these methods in terms of my adopted perspective on the validity of assessment, the unitary concept, and the theoretical criteria against which I intend to measure the validity of the assessment. It will be by using the theoretical criteria based on Messick's (1989) unitary concept that I will be able to conduct an evaluation of the validity of the assessment that encompasses both my own perspective and the claim and definitions discussed above. As construct validity is the unifying theme of the unitary concept I will need to begin by establishing how the construct, engineering competence, is being defined by BTEC, the Army, and the lecturers.

Defining the Construct - Military Engineering Competence

The aim of the course is to produce a competent military engineer with a BTEC HND so I must look at how both the Army and BTEC define what is to be taught and assessed, i.e. the construct. A key element in defining the construct is the role played by the lecturers in delivering and assessing the

construct. I will show how the lecturers are involved in deciding what should be taught and assessed.

The Army System

The Army has adopted the method of determining training requirements described as the *Systems Approach to Training (SAT)*. This system was based on a detailed analysis of the requirements of the job for which the trainee would become qualified on completion of the course. The job analysis process enabled all aspects of the job to be described in considerable detail, and emphasized the range of tasks to be performed and the standard required. The Army pamphlet No. 3 stated:

> It has been traditional to specify the training needed for a job in terms of a syllabus, a list of subject areas to be covered on a course. Experience has shown that it is much more valuable to state the training requirement in terms of behaviour rather than of knowledge, specifying first what the soldier must be able to do, rather than what he must know. (Army Training Organisation, 1993, p. 1-2)

The job analysis produced training objectives defined as statements that should specify exactly what a trainee had to be able to do, in each separate area of his job, to demonstrate that he had reached the proficiency necessary to proceed from training to work in a Field Army Unit. Training objectives were defined in terms of describing a human performance, involving a physical action, a mental process or a combination of both. The pamphlet stated that if training objectives were unambiguously stated in performance terms then all concerned in the learning process would know exactly what was expected of the student in the final assessment (Army Training Organisation, 1993). There was recognition by the Army Training Organisation of the need for clarity and transparency in terms of the domain specification. It was stated that the process was designed in this way in the belief that this ensured that the course content was relevant and complete, and so that the lecturers did not have the responsibility for deciding what to

teach. This is an interesting point as the lecturers were officially being given some leeway in deciding what and how to assess but not on what to teach. If the behavioural domain was specified clearly and unambiguously and it was clear to everyone exactly what should be assessed then the lecturers could be given some responsibility for the assessment. The problem that the Army Training Organisation appears to be ignoring is how do the lecturers decide what to assess if the domain specification results in narrow and numerous outcome statements that are, in some cases, non-assessable. They were passing on the problem to the lecturers, I have no evidence to show if this was intentional or not.

Reference to training objectives provided, according to the Army pamphlet No. 3, a clear indication of the essential knowledge, skills and attitudes that must be acquired during training. The *skills objective* was one for which learning was concerned with some physical or manipulative activity requiring movement of some muscles of the body and was not of significant relevance to the courses involved in my study as it did not include higher order skills such as problem solving. The other two were, however, extremely significant.

The *knowledge objective* encompassed both the learning of information and the processing of information. Learning information was defined as

> learning of names, labels, facts, or larger bodies of knowledge necessary for job performance and being able to recall such information.
>
> (Army Training Organisation, 1993, p. 1-7)

Learning to process information was defined as involving

> the mental skills of identifying, discriminating, classifying, and forming and using concepts, principles and rules. (*ibid*, p. 1-7)

The application of these mental skills in thinking, creating, analysing, problem solving and decision making in real situations was also added to this list. It was then stated that this type of learning was described as the acquisition of cognitive skills and was less readily observable than for physical skills. It was recognized later in the pamphlet that there was a major problem with the training objective concept when moving away from the basic skill area and into that of knowledge, as defined above. It stated that it was only possible to be quite precise both in statements of required behaviour and in the measurement of achievement for simpler levels of knowledge and procedure. There was then some recognition of the difficulty associated with defining the nature of a construct, such as engineering competence, and in assessing it.

Attitude objectives were defined as attempting to specify the sort of values and beliefs concerned with ways of behaving in the job situation, which if held by the trainee, by the time he finished training, would bring out the best work performance he could achieve. Motivation and willingness to comply were two attitudes cited. It was stated that ideas associated with attitudes were sometimes referred to as belonging to the *affective* domain. It was later acknowledged in the pamphlet that, as for the knowledge objective, it was extremely difficult to write this type of objective.

> The formulation of objectives for the affective domain is difficult to do in terms of performance, conditions, and standards, as the concepts are often vague and too open to individual interpretation.
>
> (Army Training Organisation, 1993, p. 3-7)

The SAT process appears to encompass a number of the points raised in the literature review section on competence-based assessment. The SAT system uses the notion of an integrated concept of competence as described by Hager et al (1994). The system also recognizes that the focus of the curriculum and assessment should be on the key purposes and functions of the occupation, the view expressed by Eraut (1994) and Wolf (1995, 1996) and discussed on page 31. A second issue discussed on page 31 was that of

domain specification. Whilst the Army Training Organisation recognizes the need for clarity and transparency of the domain they do not deal explicitly with the problems associated with the specification and assessment of the domain as highlighted in the literature review. The use of the three types of objective, skills, knowledge and attitude, reflects the three domains, psychomotor, cognitive and affective, defined by Imrie (1995) on page 39. The SAT process adopted by the Army Training Organisation would appear to produce an approach that should lead to the development of a sound competence-based curriculum and assessment.

The main problem with the SAT process is that it had been designed for training of basic soldiering skills and this has resulted in great difficulty in writing Army training objectives that were suitable for courses such as the Clerks of Works. This meant that this method was unlikely to produce definitive training objectives that were easily accessible, a view expressed by one of the Senior Military Lecturers.

> "Training objectives? To be quite blunt, the training objectives as exist are so broad and woolly it is almost impossible to come up with some test mechanism which would allow you to put hand on heart, and say yes, that training objective has been tested and this is the result. This is because they are just so broad it is impossible to come up with a test which would test that broadness." (ML1)

The training objectives for the Clerks of Works courses were reviewed in 1989, by one of the course lecturers, and a draft form produced that was never ratified. The current official edition was published in 1976. In 1998 the Army Training Design Team were tasked to produce the definitive version by May 2000, this deadline slipped to October 2000 and again to March 2001. Draft training objectives have been produced with only minor input from the teaching staff. There was no evidence at this stage to suggest that the production of updated training objectives will in fact change the way the course curriculum is defined and administered. There was

widespread scepticism from the lecturers interviewed about the training objectives that the analysis was likely to produce because of the nature of the Clerks of Works role, the level of technical knowledge the analysis required, and the level of expertise of those conducting the analysis.

Despite the Army view that the SAT process relieved the lecturers of the responsibility for deciding what to teach all the lecturers interviewed contradicted this, they stated that they did, in fact, decide what to teach.

> "The Royal Engineers Training Design Team comes up with a set of training objectives and we [*the lecturers*] sit down and work out what subjects are needed to clarify, to conform to those objectives. What we actually teach is up to us, but it should be dictated by the job analysis." (ML5)

> "We are given the license to decide on how best to deliver and assess the course. This is very much left to us." (ML3)

The lecturers felt the old training objectives and the new ones soon to be generated by TDT were flawed but they did all agree that the course assessment should ideally be designed around the military training objectives rather than the BTEC syllabus. I will return to this point later in the chapter. The award of the BTEC HND added another dimension to the way the construct was defined and this brings the discussion around to the BTEC system.

The BTEC System

To be awarded an HND, BTEC required each student to have passed a set number of specified units that were agreed between themselves as the awarding body and the institution delivering the course. Each BTEC unit was structured around a specific format, defined by BTEC (1986a) as: title, value, learning support time; prerequisites; summary of aims; teaching and learning strategies; assessment scheme; and, principal objectives plus indicative content/objectives section. BTEC (1986b) defined the *principal*

objectives as specifying principal learning outcomes and *indicative objectives* as emphasizing that these were more detailed and should be treated as illustrative, not prescriptive. *Indicative content* was referred to as an alternative to indicative objectives and it was defined as indicating the detailed content of a unit without employing objectives to specify it.

This approach to course and unit design was followed because BTEC (1986b) felt that a strong focus on processes and activities that provided for the application of understanding and skills was fundamental to the course design. This suggested that objective rather than content statements would be more appropriate and two more objectives were introduced, *knowledge-based* and *process-based.* The former were defined as objectives that specified the knowledge to be learned and tested and the latter as objectives that identified processes and applications. BTEC (1986b) stated that process-based objectives were more likely to encourage the development and assessment of vocationally relevant competencies than were knowledge-based objectives. BTEC associated process-based objectives with

> *application* and *invention,* but [*the process-based objectives*] may also include elements of comprehension and of *psychomotor* and *affective domains* [their italics].
>
> (BTEC, 1986b, p. 10)

BTEC considered defining these objectives in this way encouraged greater emphasis on application and invention at all unit levels. The BTEC moderator (B2) fell in line with the above doctrine and stated that he viewed thought processes as being more important than a high level of understanding of mathematics and science, which he saw as being knowledge-based subjects. He stated that he recognized the need for engineers to have an understanding of these subjects, to have acquired knowledge, and that this underpinned the processes that were learnt during the actual engineering phase of the course. He stated that as a consequence of this view, BTEC preferred to see performance-based assessments, such as projects, rather than written exams.

I have already stated that the lecturers believed that the main aim of the courses was to produce an engineer who satisfied the Army's requirement rather than to award students an HND. There was no suggestion from the lecturers or the BTEC moderator that the award of the HND enhanced the students' future performance in the job. However, the lecturers expressed the view that the BTEC HND syllabus did play a role in defining the course structure.

> "Somewhere there is a set of training objectives which are probably very old by now, about 1989 I think. Basically, it [*lesson content*] is inherited from what was here before. Certain lesson plans set out things to be covered and I try to tie those in with the BTEC units." (ML4)

> "The actual content of the course as it is at the moment is really driven by the modules [*units*] required by the HND." (ML6)

> "The BTEC modules [*units*] breakdown into more detailed objectives for particular aspects of the programme. If we take heating, then there is a breakdown there and the particular BTEC HND unit has its own objectives and refers to prerequisite units, so you can go back to those which would break things down even further into more basic material. I do refer to the BTEC objectives and I would say I am referring more and more to these because they are more specific than anything we have got on the military side." (CL10)

These comments from three lecturers, two military and one civilian ex-Clerk of Works, supported by similar comments from other lecturers, indicate that the teaching content was being guided by the objectives of the BTEC units. What would appear to be happening was that the teaching content was being structured around the BTEC HND objectives rather than the Army training

objectives despite the acknowledgement that the main aim of the course was to produce competent military engineers, not BTEC HND graduates.

So how are the requirements of the Army and BTEC brought together?

> "Historically we have managed to produce the product the Army requires. We are not taking anything out of the course we are modifying things on the course to accommodate the BTEC requirements. They have obviously got the structure there, it's a national organization with training standards and all the rest of it, which we can readily adapt to suit our needs." (ML1)

> "The BTEC HND is a secondary thing that comes from Clerk of Works training. Because the Clerk of Works training matches in part the requirement for a civilian qualification, in this case the BTEC HND, then there is a consideration of what do we cover on the Clerk of Works course that satisfies a BTEC module or part module. That is the way the overall process has been looked at in recent times. There has been a match of the requirements for a BTEC HND and what is done on the Clerk of Works course so that process has actually been gone through to satisfy the BTEC requirement." (CL10)

The source of both quotes is significant. The first quote is from one of the Senior Military Lecturers, a course director that had been in post 18 months when interviewed, and the second quote is from the civilian lecturer who was responsible for ensuring BTEC accreditation for the mechanical course. This civilian lecturer was also the ex-Clerk of Works engineer that produced the 1989 draft course training objectives. A picture is starting to emerge about the way the course modules and the content of those modules are being decided upon and hence how the construct is being defined. The BTEC units are being mapped to the modules dictated by the Army training objectives to give the course its structure and the BTEC unit objectives are

being used to define the teaching content. But there is a fundamental problem with this process, where is the link to the occupational performance, to the job analysis? The lecturers developed the academic content from the BTEC objectives, considered the Army training objectives old and dated, and yet maintained that the aim of the course was to produce competent Clerk of Works engineers and claimed this aim was being achieved. So how could this be so? The lecturers' view on this can best be expressed by the quote below from one of the Senior Military Lecturers.

> "The way that it [*the course*] has evolved is that it has come down to involved individuals delivering it and relying on a military presence in the department, serving Clerk of Works within the Corps." (ML3)

The universally held view of the lecturers, military and civilian, was that the problem with the Army training objectives was solved by the appointment of experienced Clerk of Works engineers as lecturers. The military lecturers were subject matter experts and they had the responsibility for what was taught, what was assessed, and how it was assessed. The use of professional judgement is an issue that is threaded through the theoretical discussion in Chapter 3 on the validity of assessment and competence-based assessment. Linn *et al* (1991) consider the use of subject matter experts in assessment design as being a criterion measure of the validity of assessment as do Burchell *et al* (1999). Kvale (1996) focused the role of subject matter experts as assessors in defining the valid knowledge. The profile of the individuals Hambleton and Rogers (1991) describe as being best to carry out the process of aggregation matches the profile of the lecturers. The lecturers are subject matter experts, are familiar with the content of the assessment, understand the system, and the social and political context in which the assessment is based. There is support for the use of subject matter experts in the design of the assessment however concern has been expressed over the use of their judgement. The views of Eraut (1994), Gipps (1995) and Wolf (1995) discussed on page 39 suggest that the devolution to the lecturers of the responsibility for the assessment will have serious implications for the validity, reliability and comparability of the assessment. The training of

assessors was seen by Burchell *et al*, (1999), Dunbar *et* al, (1991), Eraut, (1994), Linn, (1993), and Kane (1994) as having a critical part in overcoming problems associated with the use of the assessors' judgement about student performance.

The evidence from the first two sections shows that none of the lecturers in my study had received any training on validity, assessment design, or as assessors. There is no evidence to suggest that the lecturers set out explicitly to design assessment methods that are valid. However the fact that they are considered subject matter experts and use their professional judgement in assessing the students suggests that they may contribute to the validity of the assessment implicitly. However without processes that ensure explicit consideration of the construct and that build up consensus among the assessors, and hence the reliability of the judgement, the implicit element that rests on professional expertise will not be sufficient to ensure validity of the assessment. In fact the use of the lecturers' professional judgement may contribute little or nothing to the validity of the assessment. In the next section I will look at how the construct is assessed, I will evaluate the assessment process to explore nature of the validity of the assessment and the lecturers' contribution through the use of their professional judgement.

Assessing the Construct

This section forms the main part of my validity evaluation. The data presented and analysed will provide the means by which I will be able to evaluate the validity of the assessment against the theoretical criteria described in Chapter 3. I have included them below because they form the framework around which the data must be analysed in order to provide the evidence required to draw appropriate conclusions about the validity of the assessment.

- The construct is represented in the assessment in sufficient breadth and depth.
- Important elements of the construct have not been omitted from the assessment.

- The assessment does not introduce sources of invalidity or irrelevant variance that bias the scores or judgements.
- The way the assessment is marked reflects the manner in which domain processes combine to produce effects and is consistent with the structure of the domain about which inferences are to be drawn or predictions made.
- The results of the assessment can be interpreted to mean students that pass have demonstrated the required level of performance to be considered competent military engineers.
- The results of the assessment can be used to place graduates of the courses into jobs in the Field Army.
- A consequence of interpretation and use of the results of the assessment is that graduates go on to perform successfully in their occupational role.
- A consequence of interpretation and use of the results of the assessment is that there is a positive impact on the teaching and learning process.

I will explore what is being assessed and how it is being assessed in order to see if the assessment can be considered valid in terms of the above criteria. To collect the evidence on the criteria I will explore how the construct is being defined and assessed by focusing on data from the assessment documentation and the lecturers' interviews.

The Course Assessment Programmes

The Clerk of Works courses I studied were number 49 and 50 mechanical and number 43 and 44 electrical engineering courses. The courses were all 2 years in duration and the RSME had been running the courses and thus producing Clerk of Works engineers for the Field Army for 50 years. The content of courses in terms of module headings comes from two complementary sources, the course timetables and the course assessment programmes. The course timetables, attached as Appendix F, show the headings, the order, and the duration of each module. The course assessment programmes show the modules by phase, how they are assessed and the marks awarded for each method of assessment. In terms of data for my

study it is the course assessment programmes that provide, as I will show, important data and it is to these I now turn.

The courses were divided into three phases, foundation, allied, and engineering specific. The assessment programmes gave the headings for each module in each phase. The fundamental difference between the two assessment strategies was that the mechanical lecturers had taken the time to clearly indicate the associated BTEC units and the range of assessment methods used. However, on discussion with the electrical lecturers it became evident that the electrical modules were linked to the associated BTEC units and that the range of assessment methods used matched those for the mechanical course, i.e. assignments, examinations, and projects that incorporated presentations. I have added three columns to the programmes showing the mark allocated to the assessment method, the percentage of the total overall mark allocated to each assessment method, and the associated BTEC units. The electrical course assessment programme is shown in Table 5.1 below.

	Subject	Assessment	Mark	%	BTEC Units
Foundation Modules					
1	Mathematics 1 & 3	5 Exams	200	8.40	Maths N/H
2	Mathematics 2	2 Exams	50	2.10	
3	Computing	4 Assignments	25	1.05	
4	Engineering Science	3 Exams, 14 Labs	50	2.10	Applied Science Services, Materials
	Rounded Total % Overall			**14%**	
Allied Modules					
5	General Mechanical Engineering	2 Exams	100	4.20	
6	General Construction Engineering	2 Exams	25	1.05	
7	Engineering Drawing	1 Project	25	1.05	Associated Services Systems
8	Site Safety Supervisor Course	1 Exam	25	1.05	
9	Survey	1 Project	10	0.40	

10	Engineering Management	2 Exams	100	4.20	Tendering, Estimating & Contract Management
11	Authorised Person Course	1 Exam 1 Practical	25	1.05	
12	SNCO Course	4 Exams, 4 Practical Exercises	100	4.20	
13	Ex Leading Edge	5 Leadership exercises	25	1.05	
	Rounded Total % Overall			**18%**	
Electrical Engineering Modules					
14	DC Principles	2 Exams, 6 Labs	85 15	3.60 0.60	Electrical and Electronic Principles
15	AC Principles	2 Exams, 6 Labs	85 15	3.60 0.60	Electrical and Electronic Principles
16	AC Polyphase	2 Exams, 6 Labs	85 15	3.60 0.60	Electrical and Electronic Principles
17	Electrical Machines	2 Exams, 6 Labs	85 15	3.60 0.60	Electricity & Lighting
18	Electronics and Control Engineering	2 Exams, 1 Project	30 70	1.20 2.90	Building Services Control Systems; Electronic Instrumentation; Services Controls Technology
19	Illumination	1 Exam, 3 Labs	85 15	3.60 0.60	Electricity & Lighting, Associated Services Systems
20	Illumination Project (Syndicated)	1 Project	50	2.10	
21	Installations and Regulations	1 Exam, 3 Labs	85 15	3.60 0.60	Electrical Installations, Electrical Plant
22	Installations and Regulations Project	1 Project	100	4.20	
23	Transmission and Distribution	1 Exam, 3 Labs	85 15	3.60 0.60	Electrical Installations, Electrical Plant, Associated Services Systems
24	Transmission and Distribution Project	1 Project	100	4.20	
25	Switchgear and	1 Exam,	85	3.60	

	Protection	3 Labs	15	0.60	
26	Switchgear and Protection Project	1 Project	100	4.20	Electrical Plant
27	Temporary Power Station Briefing (Syndicated)	1 Presentation	25	1.05	
28	Industrial Attachment	Mentor Report	100	4.2	
29	Integrated Final Project (Syndicated)	1 Project	250	10.45	Administration, Integrated Project Design, Electrical Installation, Electrical Plant, Electricity & Lighting, Associated Services Systems
	Rounded Total % Overall			**68%**	

TABLE 5.1. The Electrical Course Assessment Programme.

Table 5.2 below details the assessment programme for the mechanical course in the same format as above.

	Subject	**Assessment**	**Mark**	**%**	**BTEC Units**
Foundation Modules					
1	Physics	3 Exams, 6 Labs	10 5	0.75 0.40	
2	Mathematics	5 Exams	30	2.26	Maths (N/H)
3	Higher Mathematics	2 Exams	15	1.13	Maths (H)
4	Engineering Science	2 Exams, 8 Labs	60 10	4.50 0.75	Applied Science Services, Materials
	Rounded Total % Overall			**10%**	
Allied Modules					
5	Site Safety Supervisors Course	1 Exam	5	0.40	
6	Engineering Drawing	1 Project	25	1.90	Associated Services Systems
7	Survey	1 Project	0	0	
8	SNCO Course	4 Exams, 4 Practical Exercises	0	0	
9	Ex Leading Edge	5 Leadership	0	0	

		Exercises			
10	General Electrical Engineering	2 Exams, 6 Labs	30	2.26	
11	Engineering Management & Contract Procedures	2 Exams	25	1.90	Tendering, Estimating & Contract Management
	Rounded Total % Overall			**6%**	
Mechanical Engineering Modules					
12	Climatology	1 Presentation	25	1.90	Climatology & Physiology
13	Thermodynamics	5 Assignments, 2 Exams	5 40	0.40 3.00	Thermofluids
14	Heating & Hot Water	5 Assignments, 2 Exams, 1 Project	25 100 25	1.90 7.52 1.90	Associated Services Systems; Building Services Control Systems; Air Conditioning; Heating; Thermofluids
15	Refrigeration	5 Assignments, 2 Exams, 1 Project	30 100 20	2.26 7.52 1.50	Associated Services Systems; Air Conditioning; Thermofluids
16	Air Conditioning	5 Assignments, 2 Exams, 1 Project	20 100 30	1.50 7.52 2.26	Associated Services Systems; Building Services Control Systems; Air Conditioning; Thermofluids
17	Steam & Boilerhouse Technology	4 Assignments, 2 Exams	20 80	1.50 6.00	Associated Services Systems; Building Services Control Systems; Air Conditioning; Heating
18	Water Supply & Hydraulics	4 Assignments, 2 Exams, 2 Projects	20 100 30	1.50 7.52 2.26	Associated Services Systems; Thermofluids; Hydraulics
19	Engines	2 Exams	60	4.50	Thermofluids
20	Control Engineering	2 Exams 1 Project	10 25	0.75 1.90	Building Services Control Systems; Electronic Instrumentation; Services Controls Technology
21	Industrial Attachment	Mentor Report	0	0	

22	Integrated Final Design Project	1 Project	250	18.80	Administration; Integrated Project Design; Tendering, Estimating & Contract Management; Air Conditioning; Heating; Associated Services Systems
	Rounded Total % Overall			**84%**	

TABLE 5.2. The Mechanical Course Assessment Programme.

There are a number of important elements contained in the data presented in Tables 5.1 and 5.2. The first point is that the assessment programmes and the course timetables support the earlier discussion on how the BTEC units have been integrated into what was a course designed fundamentally to assess occupational competency. This process of linking training objectives and BTEC units resulted in the use of a spreadsheet system for recording student results. Both the electrical and mechanical courses used the system, listing module heading against BTEC units and distributing marks accordingly. This system was controlled and administered by the electrical and mechanical civilian lecturers (CL5, CL10) and checked by the BTEC Moderator who reported that he was satisfied the system met the BTEC requirement (B2). The system was not easy to understand but it did create a final sheet that showed the results in terms of the military modules and the BTEC units. Mixing them in this way however raised the issue of how was the construct, and in particular the performance criteria, being defined to enable the division and allocation of marks from the assessment to the BTEC units.

An understanding of how the construct military engineering competence is being defined by the lecturers emerges from the titles of the phases, the module headings and the allocation of marks to the assessment methods for each module. The foundation phase was the academic introduction to the courses carried out within the Science and Computer Branch and accounted for 14% and 10% of the overall marks for the electrical and mechanical courses respectively. The Science and Computer lecturers who delivered this phase of the training did not distinguish between the electrical and mechanical courses in their interviews and talked in general terms

applicable to both courses (CL15, CL16, ML17). This phase covered the subject domains of mathematics, computing and engineering science. Each of the subject domains had one or more textbooks associated with it. The textbooks were all modern, well-established editions of standard college texts that governed what was taught and assessed during the foundation phase. The students were set regular homework requiring work to be completed from the exercises in the textbooks.

The allied modules focused on a small number of subjects related to both courses. The general engineering module, i.e. the mechanical module for the electrical students and vice versa, was seen as a key element of this phase because there was an Army requirement for certain jobs to be interchangeable between the electrical and mechanical engineers. The other key modules for both courses in this phase were the engineering management and engineering drawing modules. These three modules accounted for approximately 11% out of 14% of the overall mark allocated to this phase for the electrical course and effectively all 6% of the overall mark allocated to this phase for the mechanical course.

The engineering phase of the courses contained the electrical and mechanical engineering modules. The assessment programmes show that this phase was weighted the highest, 68% of the overall mark for the electrical course and 84% for the mechanical course. All subject modules had at least one or more textbooks, chosen by the lecturers as the basis for the module content and adopted because the textbooks were standard engineering educational texts published for use on college courses such as the BTEC HND. This phase included the majority of the subject module projects, the integrated project and the industrial attachment.

What has emerged from this data is that the course phases and what they contained mirrored Vincenti's (1993) model of the construct engineering competence discussed on pages 33-39. The foundation phase covers the theoretical tools, and the descriptive knowledge as defined in the quantitative data category. The allied and engineering phases cover fundamental design concepts, the technical specification, and the

prescriptive knowledge as defined in the quantitative data category. As the academic content of the courses was being driven by the BTEC objectives, the lecturers used established textbooks as the basis for the academic content, and the BTEC moderator accredited the courses, I feel there is sufficient evidence at this point to make the above comparison. The two categories about which no statement can be made are the last two, practical considerations and design instrumentalities. These categories relate to what Vincenti (1993) referred to as personal knowledge, ways of thinking and judgemental skills and Vincenti (1993) does not consider the classroom as the appropriate place for learning the knowledge and skills from these categories. Although these categories may be contained in the engineering phase, in the projects and the industrial attachment, there is no evidence at this point in my study to suggest this is so. What is beginning to emerge is the way the construct is being represented in the assessment and this is relevant to the first two theoretical criteria developed in the literature review. The requirement is now to go deeper into the assessment process and look at the assessment methods themselves.

The Assessment Methods

As I stated earlier this is not a study into the problem of domain specification in terms of what is being taught, it is a study into what is being assessed. I have collated the information contained in tables 5.1 and 5.2 to show the methods by which student performance had been assessed. Table 5.3 below shows the assessment methods for both courses and for each of the phases, how many of each method were used, and the percentage of the total mark allocated to each of the methods.

	IFP	MP	WE	L/A	CA	Pr	Ex	IA
Electrical Course Phases								
Foundation	0	0	10	14	4	0	0	0
Allied	0	2	12	1	0	0	9	0
Engineering	1	5	14	36	0	1	0	1
Total Number	**1**	**7**	**36**	**51**	**4**	**1**	**9**	**1**

% Total Mark	**10.45**	**19.05**	**56.88**	**5.43**	**1.05**	**1.05**	**1.89**	**4.20**
Mechanical Course Phases								
Foundation	0	0	14	14	4	0	0	0
Allied	0	2	9	6	0	0	9	0
Engineering	1	6	16	28	0	1	0	1
Total Number	**1**	**8**	**39**	**48**	**4**	**1**	**9**	**1**
% Total Mark	**18.80**	**11.32**	**56.26**	**11.72**	**1.90**	**0**	**0**	**0**

Key: IFP – Integrated Final Project MP – Module Project
WE – Written Exams L/A – Labs/Assignments
CA – Computer Assignments Pr – Presentations
Ex - Exercises IA - Industrial Attachment

TABLE 5.3. Number of and overall marks allocated to the assessment methods.

In Chapter 4 I explained that I chose to conduct a case study of the electrical and mechanical engineering courses because I considered there was parity between them. The data in Table 5.3, and from Tables 5.1 and 5.2, provide support for the decision to look at both courses in terms of the modular structure of the courses, the assessment methods and the weighting attached to the phases and the individual assessment methods. For both courses the percentage of the overall marks allocated to written examinations is approximately 56% and the percentage of the overall mark allocated to the projects is approximately 30%. The number of each method used in total and across the three phases is almost identical. Comparing the electrical and the mechanical courses based on the evidence from Tables 5.1, 5.2 and 5.3 shows that there is comparability between the electrical and the mechanical courses in terms of the assessment methods and weighting given to the methods. I will therefore discuss the data contained in Table 5.3 in general terms applicable to both courses.

Table 5.3 provides an insight into how the students are being assessed on the courses. What is of immediate concern is the fact that both courses allocated the percentage of the overall mark in the ratio of 56:30 in favour of

written examinations against projects. The aim of the course was to produce competent military engineers who can do the job of a Clerk of Works engineer and as both the Army and BTEC systems are performance based assessment systems this ratio is perhaps surprising. The data in table 5.3 shows the course directors and lecturers were using academic measures rather than simulated work based tasks to assess student performance in terms of a recorded overall mark. I discussed the correlation of academic and simulated work based measures to occupational performance on pages 32-33. Based on a number of studies concerned with these correlations, Wolf (1995) concluded that the correlation between simulated work based performance measures and occupational performance was much higher than for academic measures. The weighting of the marks for the overall assessment of the student is biased heavily towards written examinations and therefore raises a question mark against the use of the overall mark as a predictor of occupational performance. However, perhaps the ratio is not as surprising as it seems if viewed in terms of what it is possible to measure directly. The Army recognized the difficulty with writing objectives for the cognitive and affective domains. The discussion of domain specification in Chapter 3 highlights, as Wolf (1995) pointed out, that analysing competencies is not enough if they cannot be assessed directly. This evidence casts doubt over the use of the overall marks as a predictor of occupational performance. The evidence raises a question mark over the use of the results of the assessment to place graduates into jobs in the Field Army and the consequence that they will go on to perform successfully in their occupational role.

The view of the lecturers is reflected in the following comment from a Senior Military Lecturer responsible for aggregating the marks.

> "We are progressing down the line I would like to see us go and that is away from this formal very, very rigid examination system that has existed for donkeys years, probably since the war. You do a phase [*of the course*], you get examined and all that counts is the examination mark. This is fine as it makes life very easy for me to

> manage the course and do the statistics but I feel it does not do the students much good. There are enough idiots in the world who have proved they can pass exams but put them in a real life situation and they are useless. What we really need to have is a well practiced individual who has passed the exam for that particular phase say, but he has also done some design [*projects*] because at the end of the day that is what he [*as a Clerk of Works engineer*] will be doing in the field. By tilting the assessment towards that I think you will certainly build the confidence of the student so they can say oh yes I've done this before. You'll get a more rounded guy, a more confident guy at the end of it. If there are people who are experts at passing exams but are not much cop as engineers they will be found out by that approach." (ML1)

The above quote draws out a number of points. The Senior Military Lecturers had the responsibility for aggregating the marks that gave a final mark used to indicate that the student has passed the course. The suggestion above is that marks from written examinations are easy to aggregate into a final mark. Implicit in this statement is that aggregating statements of performance is not as easy. There is an acceptance of the need for the assessment of the academic content of the course but that the assessment should be *tilted towards* the use of design projects. Two reasons are given for using design projects rather than examinations to assess the students, it is the job of the Clerk of Works engineers to carry out design projects and passing examinations does not ensure successful performance as an engineer. ML1 could not present any evidence of a similar nature to the theoretical discussion on the correlation of academic and simulated work based tasks to occupational performance. He formed his view based on his own experiences from when he did the Clerk of Works course in 1988 and his subsequent work experiences as a Clerk of Works engineer.

I stated earlier that BTEC preferred to see performance-based assessments such as projects because of the need to assess process-based objectives. The

BTEC moderator considered range of assessment methods used was good and the balance was about right between them (B1). It was clear from the moderator's reports, which were reviewed from December 1996 to July 1999, that the BTEC moderator had agreed that the projects were considered important. The general comments from the reports gave an indication that the use of projects was seen as compensating for the use of written examinations. In a report dated May 99 the moderator stated that:

> "If some of the written tests are a little routine, this is more than compensated for by the major projects which call for a significant degree of application, invention and initiative in highly practical contexts. Candidates are well prepared and briefed for the assessment tasks."

So a paradox has emerged. The lecturers and the BTEC moderator see projects as being the assessment method that should be used to assess the students performance yet the aggregated mark indicates that 56% is awarded to written examinations and only 30% to projects. The results of the assessment are used to place individuals into a job as a Clerk of Works engineer and interpreted to mean that the individual is occupationally competent. The aggregated marks are heavily weighted towards academic measures and the evidence presented by Wolf (1995) suggests that the results from the projects would give a much clearer indication of occupational performance. This paradox again raises the issue of specifying and assessing the domain. The discussion in the literature review, pages 31-33, highlighted the difficulty in specifying the domain clearly and unambiguously in relation to competence-based assessment and hence to occupational performance. Although there is a clear feeling that projects should be the dominant form of assessment method the difficulty in developing assessable performance criteria for projects could be a contributing factor to the predominant use of written examinations. Written examinations are generally associated with the academic curriculum and subject domains that may be more clearly and easily defined and assessed. Wolf (1996) suggested that the use of direct periods of employment could be included to supplement or replace direct assessment and the industrial

attachment at the end of the courses matches this requirement. I will now set out to explore these issues by looking at the written examinations, the projects and the industrial attachment in more detail.

The Written Examinations

The combined number of written examinations for the electrical and mechanical courses in the *foundation phase* was 24. Analysis of the written examinations showed that there was an average of 12, a minimum of 9 and a maximum of 14 questions per examination. Questions on the written examinations had up to five parts requiring the students to explain a concept or devise an equation and then carry out related calculations. It was evident that thought had been given to the design of questions for common subjects like mathematics and physics in terms of face validity as each course had similar examinations with either an electrical or mechanical theme. A simple example of this would be transposition of formulae where electrical students were given Ohms' Law to transpose and mechanical students were given the formula for kinetic energy. In mathematics the progression of the examination papers started with a paper on basic algebra and ended with a paper on Fourier Series and Laplace Transforms. These two subjects comprised the higher mathematics module and students of both disciplines had to pass this module as it was a requirement for the award of the BTEC HND. The progression can best be represented by the questions, taken from the series of mathematics examination papers, shown in Table 5.4 below.

Mathematics Progress Test Paper 1	
Q1.	The length of a plate detail is 891 mm. Rivets are placed 45 mm apart and the distance between centres of the end rivets and the edges of the plate is 18 mm. Determine the number of rivets required. (3 marks)
Q.2	Factorise the following expression: $k^2l^2 - mnl - k^2l + mn$ (3 marks)
Mathematics Examination Paper 1	
Q.1	The height 's' in metres, of a mass thrown vertically upward in a time 't' seconds, is given by the expression:

	$s = 40\,t - 13t^2$ By solving this quadratic equation, find the times at which the mass is at an altitude of s = 25 m. (8 marks)
Mathematics Examination Paper 2	
Q.2	a) The distance 's' cm of a vibrating particle is given by the following expression where 't' is in seconds. Find the velocity when 't' = 50 ms $s = 3 \sin (3t - \pi/5)$ (6 marks) b) A missile fired from ground level rises 's' metres in 't' seconds where: $s = 75t - 12.5t^2$ Determine: 1) initial velocity 2) the time when the height of the missile is at its maximum 3) the maximum height reached 4) the velocity with which the missile hits the ground (8 marks)
Higher Mathematics Examination Paper	
Q.3	A vertical spring and mass is immersed in a viscous liquid as depicted in the diagram below (*not shown here*). A mass of 2 kg extends the spring by 4 cm. If the damping factor is 3 and is proportional to the instantaneous value of the velocity. The system can be solved by the differential equation: $M\,d^2x/dt^2 + kx + \lambda\,dx/dt = 0$ Determine: a) the value of the damping factor λ b) spring stiffness constant k c) if a mass of 3 kg is suspended and the spring initially extended 5 cm and then released (dx/dt = 0) find the subsequent motion. (25 marks)

TABLE 5.4. Sample of questions taken from the mathematics examinations.

The results of the assessment of the foundation phase were passed to the Senior Military Lecturers in the form of an overall mark for the phase, and an individual mark for each element i.e. laboratory assignments, progress tests and written examinations. The overall weighting for these assessment

methods was biased towards the written examinations, 12% out of 14% for the electrical course and 8.5% out of 10% for the mechanical course. However there was more to the results of the assessment than just the marks. The results of this phase also included associated comments on each individual student. Examples referring to different students taken from the standard written reports submitted by CL16, an ex-Clerk of Works now a civilian lecturer, are given below.

> Sgt X demonstrated a cheerful disposition, a willingness to question and a determination to learn. He put a great deal of time and effort into his work that resulted in a consistently high performance throughout the phase. Sgt X contributed in class discussion and worked well in a team. His continuous enthusiasm gave rise to an interesting course phenomena, the 'X' factor. This will appear again, no doubt.
>
> Sgt Y found this phase, and in particular the academic aspects of it, very demanding. Following a poor start in both his mathematics and engineering science, due mainly to a stressful, self-imposed demand to succeed, rather than inability to do so. He re-assessed his approach and was soon able to produce some very acceptable results.
>
> A team player, Sgt Z contributed in all classroom and laboratory activities. His ebullient personality and professionalism stood him in high regard with his peers.
>
> Sgt U contributed well in class and worked well in a team. He carries forward an excellent resolve and ability into the next phase.
>
> Sgt V has developed his self-confidence, works well in a team and contributed greatly during the practical laboratory assignments.

> Sgt W has performed exceptionally well throughout the phase and displayed considerable ability in the field of mathematics. This, combined with a natural engineering flair, makes as good as a foundation to his chosen career as is possible to obtain.

None of the formal assessment methods measured such things as resolve, self-confidence, enthusiasm or engineering flair. The first, third, fourth and fifth comments above did give an indication of how this was done, and the interview with the lecturer (CL16) who wrote them confirmed that he formed his judgement during the daily interaction with the students in class and in the laboratory. He was asked why attitudes and behaviour were considered an important part of the assessment of the foundation phase.

> "There must be an assessment of the individuals attitude, determination. He has to apply himself as this is a very demanding phase. If that determination isn't there then traditionally you will find the results drop off, interest will drop off, because the potential is not there. That is a significant indicator that the technologies they will be subjected to will be beyond their grasp. The results generated from this foundation phase, reaching a certain mark, is certainly an indicator of their ability to go on and pass the course." (CL16)

What was evident was that this phase was also thought to give a clear prediction of future success in passing the course (ML1, ML3, ML4, ML6, ML9, CL10, CL15, CL16, ML17). The comment below from one of the Senior Military Lecturers expresses the general view.

> "Traditionally, the marks attained in the foundation phase are a good indicator of future performance and the ability to pass the course as a whole. It is very rare for someone who has done well to fail and those who struggle continue to do so throughout the rest of the course." (ML1)

There are three points that arise from this data that relate to the discussion on the interpretation of the results of assessment contained in Chapter 3. The first point is that the lecturers clearly considered that predictions about the students' ability to pass the course could be made from the results of the assessment of the foundation phase. The main assessed elements of the foundation phase are mathematics and engineering science. The lecturers' view supports the findings of the study by Williams and Boreham (1991) who concluded that there was evidence to suggest that student performance in mathematics, physics, and engineering drawing examinations could provide a reasonable prediction of success in terms of passing engineering courses. The second point is that at no time did the lecturers indicate that they considered these results could be used to predict future vocational success. As this is an academic phase of the course and assessment consists mainly of written examinations, the studies critiqued by Wolf (1995) and discussed on page 32 support the view that the results of this phase would not give a good indication of future occupational success.

The third point is concerned with the issue of students failing the academic assessment. Wood (1991), supported by Gipps (1995), discuss the fact that there is no way of knowing how students that fail would have performed in the job i.e. in my study as Clerk of Works engineers. During my study two students failed the courses. Both performed poorly on the assessment of the foundation phase, one in fact being Sgt Y above, and both were removed during the early modules in the engineering phase, one failing the Alternating and Direct Current Theory examinations and Sgt Y failing the thermodynamics examinations. There was evidence then that performance in the mathematics and science examinations did give a prediction of the students ability to pass the course, or more precisely, to pass the rest of the written examinations on the course. Neither of these two students had carried out any design projects and based on the previous discussion about the lecturers views on written examinations and projects in assessing student performance in relation to the job should there have been doubts about failing them at this point on the course? The dilemma is that the lecturers could not have known whether either or both of the failed students could have gone on to become competent Clerk of Works engineers.

In the *allied and engineering phases* there were a combined total of 51 written examinations. My analysis of the examinations revealed that the average number of questions was 6 per examination with up to 5 parts to a question. As in the foundation phase, the lecturers used textbooks and the BTEC moderator was happy that the content of the examinations met with BTEC requirements. I compared the contents of the electrical machines, illumination, installations, transmission, and switchgear examinations from the electrical engineering phase, and the heating and hot water, refrigeration, air conditioning, and water supply examinations from the mechanical engineering phase against the contents of the textbook for each of the modules. My analysis showed that the content themes of all of the questions on all the examinations could be found in the textbooks. Table 5.5 below shows the comparison for the heating and hot water module examination paper A.

Heating and Hot Water Module Examination Paper A		**Faber & Kell's Heating & Air Conditioning of Buildings Martin & Oughton (1997, 5th Ed)**
Question (Marks)	**Topic**	**Textbook Reference**
1.a (6)	Comfort Conditions - 3 Main Parameters	Page 16
1.b (8)	Temperatures - Dry Resultant, Internal Air, Environmental, Mean Radiant	Pages 9-11, 48
2.a (12)	Rate of Heat Transfer	Pages 172-4
2.b (9)	Interface Temperature	Pages 33-40
3.a (4)	Explain - Indirect Hot Water Supply System	Pages 139-72
3.b (8)	Diagram - Indirect Hot Water Supply System	Pages 139-72
3.c (3)	Secondary Return	Pages 588-93

4.a	(3)	Optimisers 3 Main Functions	Pages 602-16
4.b	(8)	Diagram - Optimers and Compensator Set-up	Pages 602-16
5.a	(4)	Progressive Cooling & Index Circuit	Pages 396-9 Pages 147-53
5.b	(2)	Direct Hot Water Supply System - Disadvantages	Pages 94-104
5.c	(6)	Indirect Hot Water System - Gravity Circulation	Pages 243-6
6.	(8)	Centrifugal Pumps - Characteristic Curves for Pump and System	Pages 201-7

TABLE 5.5. Heating and Hot Water Examination Contents.

Having collected data from the written examinations, the analysis confirmed that the content of the examination related to the contents of the textbook. It did not provide evidence of what was taught in class and I did not attempt to sample the classroom teaching environment. I found the analysis of the examinations difficult because I was not a subject matter expert in many of the subject areas. I will use examples of questions taken from the water supply module examination papers and comments made by the lecturer responsible for delivering and assessing that module to discuss what was being assessed generally by the examinations.

Paper A Question 1.

a. State the maximum practical suction lift that can be expected for a velocity type pump and a reciprocating pump when operating at sea level pumping fresh water. (5 marks)

b. What factors affect the suction lift of a centrifugal pump and why? (5 marks)

c. Explain why a single stage centrifugal pump is effected by end thrust. (5 marks)

d. Explain with the aid of sketches, two methods of balancing end thrust in a single stage centrifugal pump.

(5 marks)

Paper B Question 3.

A pipeline is used to connect two reservoirs with a difference of 10 m between the water levels. The pipeline consists of a single 600 mm diameter pipe 3500 m long, feeding a junction from which three 200 mm diameter pipes, operating in parallel, traverse a further 2500 m to the lower reservoir.

If the friction factor of all pipes is 0.01 establish the total discharge into the lower reservoir in litres per second.

Neglect velocity head. (15 marks)

The lecturer was looking for evidence of the student's knowledge of the device, the water pump, and the procedures for carrying out the pressure calculation that linked the pump to the pipeline, both of which defined the water supply system. All the examination papers followed a similar format assessing the students knowledge of the procedures used in the design calculations and the operational principles of the mechanical or electrical devices used in the design. In the case of the water supply module the military lecturer described the form of the examinations, thus:

> "There are two 3-hour exams. One mostly written and one mostly calculations. But saying that, some of the written one has got a 'what do you understand by this' and then a small calculation to prove the point, and on the calculation ones there is an 'explain this' and then calculate this. So what we are trying to prove by that is that they have taken in the knowledge that has been put across to them and they have learnt what is of practical use to them. Deriving the formula is not of practical use to anybody. Some people like those sorts of questions. I would rather see them

> doing pressure calculations for a pipe, what size pump do you need for this. Lots of 'what do you understand by this' and 'how would you have overcome that' rather than 'derive this' and 'what's the flow of over this V-notch' sort of thing." (ML8)

The military lecturer was designing the written examinations to assess objectives from the curriculum content and the cognitive processes as defined by Haertel's (1985) model of the construct. There is evidence that elements of Vincenti's (1993) first four categories were being assessed through the use of written examinations in the foundation and the engineering phases. The assessment was focused on the academic nature of the course rather than the requirement to measure occupational performance. Although there is some evidence of cognitive skills being assessed there is no evidence of assessment of behavioural outcomes or the affective domain. In terms of interpreting and using the results of the assessment as a predictor of occupational success the emphasis on written examinations and their low correlation with occupational performance raises doubt over the use of the results of the assessment as a good predictor. There is still a clear requirement to assess the two domains, cognitive and affective, with emphasis, as Imrie (1995) states, on understanding and grasping concepts and being able to apply them creatively. At this point in my study there is little evidence of the assessment of competence in terms of the ability to use knowledge, skills and attitudes in the successful completion of engineering tasks, as discussed in Chapter 3. The quote from CL16 hints at where this may begin to happen.

> "I feel the way we develop skills, like communication and analytical skills, is through the introduction of short presentations and various tutorials within the class where we chat and talk about things, problems. Gradually working up to the point when we introduce the major project and are able to sit there and almost argue with them about their suggested solutions. They are actually able then at a later stage to be able to counter argue and

defend their solution. Whereas we know early on in the course they would probably just look at us and say well you are right and we must be wrong. Certainly later on that does not happen and they are prepared to stand their ground which is what we are trying to achieve, although at the same time we want them to be able to listen, to accept other proposals but then sensibly accept or dismiss them." (ML9)

There appeared from the comment above that the lecturers considered communication skills and the ability to discuss design solutions, listen to alternatives and make decisions to be desired outcomes. There is an indication that the development of these skills began at an early stage of the course and continued throughout but that the way they were assessed was in the final integrated design project. The design projects will now be discussed in detail.

The Design Projects

The design projects all required the students to carry out the design process appropriate to the subject module being assessed. The projects were simulated work-based tasks designed, administered and assessed by the lecturer that delivered the module. There were 7 projects used on the electrical and 8 on the mechanical course. The marking guides for all the projects showed that there were a considerable number of elements being assessed, the marks awarded to them were very small, and they focused on output in terms of the written results rather than on student performance during the project design process. For example in the electrical installation project marking guide all the marks related to the engineering procedures and the final design solution except for the 5% allocated to individual contribution, see Table 5.6.

Section		Value	Mark
Design Information Pack		**55**	
Design Data and Assumptions		**12**	
	Set out of data	4	8

	Assumption justification	8	16
Engineering Calculations		**12**	
	Contents sheet	4	8
	References	8	16
Sketch Drawings		**12**	
	Clarity	4	8
	Labels	8	16
Computer Design Output		**12**	
	Contents sheet	4	8
	Cross-reference	8	16
Usefulness as a Reference		**7**	14
Technical Content		**40**	
	Fire alarm system	**5**	10
	Security system	**3**	6
	PA system assessment	**2**	4
	Lightning protection and internal earthing	**5**	10
	Office underfloor wiring system	**3**	6
	Schedule of installed / diversified loads	**5**	10
	Schedule of external influences	**2**	4
	Detailed materials/cost schedules	**5**	10
	Drawings	**5**	10
	Manpower, time, tools and equipment	**5**	10
Individual Contribution		**5**	10
	Total:	**100**	200

TABLE 5.6. Electrical installation design project marking guide.

There was no indication in any of the marking guides of formal assessment being made with regard to how the students carried out the process, what they did to produce a design report. It was also not clear what the marking criteria were, except that there were many of them and that they were linked to the contents of the design report. The water supply project pack at Appendix G with its 10 page marking guide showing the available 2272

marks, is a particularly good example of narrow and numerous criteria being generated to the point of each criteria becoming meaningless.

The marking guide for the mechanical integrated final project was divided into three sections, the feasibility and initial design briefing, the written report, and the final design presentation. Single marks of 15% and 20% were allocated to the feasibility presentation and the final presentation of the design respectively. The written report was assessed using a series of questions for which marks were awarded based on how the question could be answered covering three areas, the swimming pool zone, the gymnasium zone and the office zone. There were a total of 11 main questions and 20 secondary questions, and Table 5.7 shows two main questions with the associated secondary questions taken from the marking guide.

Criteria	Marks Available
Has a service been designed to control the summer time temperature?	Yes (5) No (0)
Design concept?	Good (5), Sufficient (3), Flawed (0)
Design specification?	Good (5), Sufficient (3), Flawed (0)
Design drawings?	Good (5), Sufficient (3), Flawed (0)
Control of system? Adequate/clear/possible?	Good (5), Sufficient (3), Flawed (0)
Has a service been designed to control the summer time temperature?	Yes (5) No (0)
Design concept?	Good (5), Sufficient (3), Flawed (0)
Design specification?	Good (5), Sufficient (3), Flawed (0)
Design drawings?	Good (5), Sufficient (3), Flawed (0)
Control of system? Adequate/clear/possible?	Good (5), Sufficient (3), Flawed (0)

TABLE 5.7. Extract from the mechanical integrated final project marking guide.

The marking guide for the electrical integrated final project followed a slightly different approach to its mechanical counterpart in that it consisted of a series of sheets, each assigned to a particular topic area, and a master sheet that collated them. The electrical system had not developed any marking guides other than those with single headings. Table 5.8 shows the master sheet. The students were not shown the marking guides for either the electrical or mechanical integrated final design projects.

Section	Max Mark
Report Writing	100
Civil Works schedule	50
Power stations/supplies	100
Transmission & Distribution, including grading & problems	100
Street, security, floodlighting	100
Lightning Protection	50
PA system, station sirens	50
Load assessments	25
AGL details	100
Outline design heating, ventilation, hot water	50
BFIs and fuel storage	50
Cookhouse	50
Borehole pumps	25
Supply to 800 man camp	25
Hanger lighting	25
Individual Assessment	100
Totals:	1000

TABLE 5.8. Electrical integrated final project marking guide.

The above headings were broken down into sub headings that were again single statements. The breakdown for the individual assessment is shown in Table 5.9.

Assessment	Max Mark
Project awareness	15
Communication skill	15
Attitude	15
Report writing	20
Mark for actual design carried out	35
Total:	100

TABLE 5.9. Electrical integrated final project individual assessment guide.

The integrated final project marking guides showed that 35% of the marks allocated in the mechanical project and 10% for the electrical project related to performance criteria assessing behavioural outcomes and the affective domain. The marks allocated equated to 6.8% and 1.45% of the aggregated overall mark for the mechanical and electrical courses respectively. The evidence of what is being assessed as the construct remains firmly focused on Vincenti's (1993) first 4 categories except for the elements identified in the projects as the presentations and the individual assessment. I will now look at how and what was being assessed by the lecturers in these elements. What the lecturers were assessing is demonstrated in the quotes below from four military lecturers.

> "We look at the project as a whole, we are looking for a coherent and sensible plan around the problems they have been set, bearing in mind that whatever they produce would have to go, in reality, to some higher command. They are experiencing a learning curve as they go through every project. The projects are all design based and we are looking for someone who can give a well balanced, well thought out design. Someone who can work out a design, then produce arguments and assumptions to reinforce and support that design." (ML2)

> "We are looking for them to be able to analyse a problem i.e. this building needs some sort of environmental control, they determine the level of control they feel should be

necessary and they then would propose how they would achieve this. It is a technical solution, yes, but at the same time it's their being able to tell us what that technical solution is and then being able to have confidence in it."

(ML6)

"It [*the integrated final project*] tests, among other things, their ability to read into the project. It assesses their abilities to glean information, technical information that we have given them already, and their communication skills." (ML7)

"Deliberately we leave things out or we leave them vague [*in the project brief*]. What we are looking at is their ability to come back and ask for more information once they have looked at it in more detail." (ML1)

The lecturers were looking for the ability to analyse a problem, to justify their solutions with confidence, glean information, and communicate ideas. The lecturers all had a view on what they were looking for in terms of behavioural outcomes but how were they assessed and what was the required standard?

"They have to come and give us a brief as a team and part of the brief is that all members of the team must have their input, the whole thing is put across to us as a designer client. They want to show us they are the best team to do this work." (ML6)

"Generally three of us are in the meeting with the students and then between the three of us, once they have actually left the room, we will discuss what we thought were the particular strong points and what were their weaker points." (ML7)

> "The biggest test is whether the report could be handed over to a constructing unit to carry out the task. That is the biggest test and then more critically we are looking at the actual design itself. We are not just looking at what is produced here. We will have interviews with the student or the team as whole. We will look at them as a design team. We will have talks on a regular basis, generally a weekly basis." (ML9)

There was no feasibility presentation included as part of the electrical project, what took place was a series of interviews and correspondence between the student teams and the electrical Senior Military Lecturer, usually accompanied by one of the lecturers. All the interviews were conducted by the Senior Military Lecturer during the middle two weeks of the project and were dependent on student progress. Although the Senior Military Lecturer ensured all students were interviewed to discuss their input and progress, there was no apparent structure in terms of timings of the interviews or their content. Unfortunately I was unable to observe these interactions and therefore have no evidence of what took place during this process.

As part of the integrated final project for both courses the students gave a final design presentation and at the same time they submitted their team's final design report. The final design presentation took a similar format to the feasibility presentation and the marks were discussed and awarded after the presentation. I observed this process and found that the military head of department was also present, thus making an assessment team of four lecturers. No formal marking schedule was used. Each of the students presented an element of the design solution and answered questions on the solution presented. The subsequent discussion between the lecturers centred on the feasibility of the design solution presented to them in each of the areas and whether the complete design solution was a solution they thought would work. At no time were any behavioural outcomes or attitudes discussed, questions like 'how did he argue his case' were not addressed. There were no performance criteria explicitly stated for these behavioural

outcomes and it is not surprising then that the process by which they were supposed to be measured did not do so formally. There was evidence to suggest the way the students' behaviour was assessed throughout the course and that through the use of the lecturers as the assessors compensation for the students was occurring. An example of how the lecturers were compensating the students is given in the part of the transcript with CL10.

> DB: "If I were to ask you who was the best at this or that would you be able to tell me?"
> CL10: "I think yes, certainly. I could discuss it and give an opinion of each student."
> DB: "And how are you developing that view of the student?"
> CL10: "By performance in the classroom, by students themselves having to stand up and present something, just meeting the students generally, that's how I form an opinion."
> DB: "Does this view influence your interpretation and assessment of projects?"
> CL10: "Perhaps unconsciously. I suppose in some respects yes. If I am reading something written by a student I feel has not got the grasp of writing it in a logical way but I think well okay, I know this guy, and the way he thinks and speaks, but he can't write it down then this perhaps influences the way I assess his work. I find it difficult to answer honestly, except that I do look at a student and think is he a confident, go ahead guy."

The project marking guides show that the emphasis of the recorded marks was firmly placed on the design solution presented in the design report and the final presentation. What was being assessed was the ability to produce a technically correct design and an accompanying report. However, the marking guides contained many criteria to assess and I found no evidence of the standard against which the criteria should be measured. The problem surfacing yet again is that of domain specification, the lecturers have ended

up with narrow and numerous criteria that appear to be easy to measure and in danger of becoming meaningless by allowing the lecturers to compensate for the students when allocating marks to the criteria.

The lecturers acknowledged that the courses should be designed to meet the Army training objectives because the students that graduate will join the Field Army as Clerk of Works engineers, yet the assessment has focused on academic measures rather than simulated work based methods i.e. the design projects. BTEC considered that projects were essential in assessing process-based objectives and the BTEC moderator was pleased to see the emphasis that was placed on the use of design projects to assess the students on the courses. As the simulated work based assessment used to assess the students on both courses, there was a fundamental problem with the marking criteria and with the interpretation of the results. The design projects should be assessing criteria or outcomes from the cognitive, affective and behavioural domains related to the occupation i.e. military engineering. The evidence shows that the formal assessment of the projects contains only a small element associated with behavioural outcomes and in fact the students are being assessed against these outcomes throughout the course and in a number of ways. The claim was made by the lecturers that they produce what the Army wants, individuals that can do the job of a Clerk of Works. But is this really the case and is the claim being made for the results of the formal assessment?

> "DB: Are the design projects actually going to assess whether they can do the job or not?
>
> ML8: Yes, to a certain extent. But it is something that they are not going to be able to do properly. They can do the project now, get a mark for it, come back in three years time when they have done it for real and do the project and they would do it a totally different way. Because their knowledge is a thing that you can't give them, okay two years [*on the course*] is a long time, but you can't get them to go out and practice all the things."

Another example is given below.

> "DB: They do the two year course, they get assessed and then go out and do the job. How well do you think the course assesses them in terms of going out and being a good engineer?
>
> CL10: I find that difficult to answer. I feel that the assessment on the course assesses that they have retained knowledge of what they have just been taught and I think the course as a whole teaches them sufficient knowledge to be able to go out there with some confidence and approach engineering tasks. I think it does it well but the big thing that brings them on to being a good engineer is not just the course but experience that they gain once they have left the course. I don't think they can gain enough experience in the particular subject matter that they cover, they can gain the knowledge but the experience is not going to come until they begin on the attachment. I would say, and I am perhaps really talking about myself now, that when I left the course many years ago if someone asked me if I were a good engineer I couldn't have even said I was a confident engineer on leaving the course. But once you get into doing particular tasks, real tasks on the ground, building your experience then hopefully you become a good engineer. The course as a whole is going to be a foundation to the guy becoming a good engineer."

There is a major implication from the above statements. The claim being made for the assessment was that the students that pass the assessment should be capable of developing into competent Clerk of Works engineers as they gain experience. There is a clear link to Vincenti's (1993) view that personal knowledge is an essential part of being an engineer and this knowledge cannot be taught in the classroom it can only be learned through practical experience. In the light of Vincenti's (1993) and Glaser's (1993)

distinction between novices and experts, the results of the assessment produces novices lacking the personal knowledge of the more experienced, expert, Clerk of Works engineers. However the Field Army expects the graduates to be competent military engineers on completion of the training courses so there is evidence of a possible mis-match occurring between expectations of and what can be reasonably achieved through training. The expectations of the Field Army concerning the competence of the graduates should encompass Eraut's (1994) two dimensions, scope and quality discussed on page 30, and the notion of competence based on a model of progression and lifelong professional learning. There is one element on the course that does involve practical experience, the industrial attachment, and this will now be discussed.

The Industrial Attachment

The purpose of the six-week industrial attachment at the end of the course was to give the students the opportunity to work in the engineering community. An opportunity to learn what Vincenti (1993) referred to as judgemental skills such as insight, imagination and intuition, which are mostly tacit and learned through practical experience. Although knowing how to use these skills, this knowledge, becomes an individual thing based on the wide range of experience that will be different for each Clerk of Works engineer working in the Army, this phase was seen as learning to do it for real. The lecturers and the students considered this to be the key element of the course, and it is supported by Wolf's (1996) view that direct assessment could be replaced by experience to help combat the problem of assessing objectives and outcomes. Vincenti (1993) considered the categories practical considerations and design instrumentalities could not be taught in class and learned only through experience. The following quotes reflect the importance the lecturers placed on the attachment.

> "The attachment puts them [*the students*] into a real life working environment to increase their confidence through knowing that what they have done on the course is

actually or can actually be used in a real life situation."

(CL10)

"The attachment is important because they have only ever done it in the classroom and on paper before, and they haven't got the confidence that they are right. They have designed a pipe over a hill and into a water treatment, but if it doesn't work it doesn't matter. It's on paper. And they haven't got the experience to know that it will work. Then if it was on paper, did it work, or it might have done. It doesn't help your confidence. If you go out there and somebody gives you a job and says "make this work" and [*you*] apply the same theories, then it works, then the confidence is there, that's the thing they need which you can't give on paper." (ML7)

"I think it [*the attachment*] is very important. I would class it as an important stepping stone from the formal learning procedure they have just gone through to putting it into practice where they are literally going to be put on the spot everyday of the week. It [*the attachment*] gives them a chance to practice what they learnt on the course prior to their military posting. Again it is a boost in confidence because they do then tend to know that they know more than they think they do." (ML9)

There was no formal assessment of this phase. The students were assessed by whomever they worked for during the phase, and one of the military lecturers visited them once during the attachment.

"There is not much assessment done other than a visit by the staff to the engineer [*civilian mentor*] employing the student and discussing with him how the student is getting on." (ML9)

The students then are given the opportunity to develop as engineers through practice in the civilian engineering community. The assessment of this phase was based on the mentors' view of how the student performed and was weighted as 4.2% of the overall mark on the electrical course and 0% on the mechanical course. The purpose of the attachment was not really then to assess the students' performance but to give them an opportunity to carry out engineering tasks in the real world which would support the views put forward by Wolf (1996) and Vincenti (1993). The statements from the lecturers suggested the attachment was very important in terms of developing the students but this is not really reflected in the overall marks, the results, of the assessment. The difficulty faced by both the mentors and the lecturers is again that of specifying the domain and the performance criteria that should be assessed on the attachment. This problem is further exacerbated by the fact that the experiences were different for each student, most arranged their own industrial sponsor, and that the attachments were carried out in civilian and not military environments.

Other Evidence

Introduction

The data presented thus far has referred directly to the assessment process and the lecturers' involvement in that process. The purpose of this section is to present the views of the students, serving Clerks of Works and their employing officers. I have included comments from the BTEC Moderator in the earlier sections. This section is a discussion of other evidence against which the assessment and the judgements made by the lecturers will be evaluated, a point seen as important by Messick (1989) because the fallibility of expert judgement should be evaluated on the basis of other evidence. I will begin this section by evaluating the data from the students and then the data from the TDT (1999) Review discussed earlier.

The Students' Perception

I considered the students' views were important because they supplied other evidence about the assessment process in terms of the theoretical criteria against which the validity of the assessment is being measured. The total number of students on the courses during my study was 34. The data presented originates from the student questionnaires and the student interviews. This section will focus on the students' views on the assessment process and the comparison of their views with the data presented in the earlier sections. The emphasis will be on assessing the construct and the interpretation and use of the results of the assessment.

The students supported fully that the academic content of the curriculum was well represented in the assessment process and that written examinations and design projects assessed what they had been taught in class. They saw the written examinations as most important in assessing BTEC objectives and the design projects as most important in assessing the military objectives. The students indicated (Q1, Q2, S1, S2, S4) that they felt the purpose of all the design projects was to assess their problem-solving skills, understanding of the subject matter, and personal qualities. The set of personal qualities presented to them was chosen as a result of the initial interviews with the lecturers and from the Army and BTEC publications discussed earlier. The students were asked to indicate which qualities were being assessed formally and informally. The evidence shows that the students thought that there were elements from the cognitive and affective domains that were being assessed by the lecturers. The evidence presented earlier showed that the assessment of these qualities was not recorded formally. The results from the student responses were as follows.

	Formally		**Informally**	
	Yes	**No**	**Yes**	**No**
Working under pressure	27	7	33	1
Communication skills	31	3	27	7
Motivation	29	5	33	1

Determination	31	3	33	1
Working as part of a team	30	4	33	1
Creativity	28	6	27	7
Initiative	29	5	28	6

TABLE 5.10. Student response to question 4 from the end of course questionnaire.

Although 65% of the students indicated that they were clear about what they were required to do in terms of output for the projects, 80% indicated that the performance criteria used to assess the projects were unclear. Again the earlier evidence showed that not only were the marking guides unclear, they were not issued to the students which would suggest that the students based their responses in Table 5.10 on how they perceived the lecturers assessed their progress. This view was supported during the interviews.

> "It's not clear what they [*the lecturers*] are looking for. They mark the projects based on how they would have done it, based on their experience. I think that works okay, it seems fair." (S2)

The evidence suggests that the students were acknowledging the use of the lecturers' judgement as experienced subject matter experts and accepted that this was an integral part of the assessment process. One area the students were not happy with, and that provided the only instance of possible construct irrelevant variance, was their concern over the content of some of the written examinations. The student responses to question 3 from the end of course questionnaire indicated that 77 % of the students found some written examination questions were presented in a confusing manner and 88 % felt that the confusion could be reduced by relating them directly to engineering topics, i.e. giving them face validity.

> "Some exam questions were similar to examples we had done in class but others were not like any we had ever done before. They [*the questions*] seemed almost like trick

questions. Variations on a theme are okay but there should be limits so that we still understand the question." (S3)

The responses to question 5 from the end of course questionnaire showed that 91% of the students viewed the written examinations as being the form of assessment most likely to predict success in terms of completing the course. Only 15 % viewed written examinations as most likely to predict future job performance. The students agreed with the lecturers' view that the foundation phase contained the important underpinning knowledge and that the assessment of this phase did predict future success but only for passing the rest of the course, not success in the job. The students typically commented that:

> "The maths and science, particularly the maths, will be used in other subjects. So yes I think the results from the foundation phase are quite accurate in predicting the ability to pass the course but not to do the job." (S1)

> "This phase tested our ability to absorb knowledge and to cope with the pressures of the course. Success during this phase will lead to success on the course as a whole." (S4)

The students were clear about the interpretation and use of the results of the foundation phase assessment and their views support the work of Williams and Boreham (1971). They were not quite so sure about the results of the assessment being an indicator of future occupational success. All, 100%, of the students viewed the design projects as most likely to predict future job performance against 65% who considered them most likely to predict success in terms of completing the course. The reason given for the difference in opinion can be summed up by the following quote.

> "Written exams are testing what I can remember, memory. They [*the exams*] are not telling how well I would do the job, what I am able to do. The projects do that." (S3)

The students' views clearly fell in line with the earlier discussions in this Chapter and Chapter 3 on the correlation between assessment methods and occupational performance. The design projects were considered by the students to be the most likely to predict how good they would perform as an engineer because *when we get out into the real world that is what we are going to be doing* (S4). They were concerned with producing a good solution because this was what they believed was assessed but they felt this was difficult because they did not have the experience to know if their solution would work or not, they lacked personal knowledge (S1, S2, S4). Although comments about the industrial attachment were not collected from the questionnaire, an oversight that became apparent during the analysis, the students were asked to give their views during interview.

> "Yes, I definitely feel confident after the attachment. The interaction was good and they [*the engineers*] said I could have a job there anytime." (S2)

> "The attachment makes you feel better, nobody knows how to do the job before then and the experience gives you the confidence that you can stand on your own two feet." (S4)

The students viewed the industrial attachment as an opportunity to get experience on-the-job that gave them an indication of how they might perform as Clerk of Works. Having completed the assessment process the key question was did the students think the course assessed successfully their potential to perform well in their future job. The students were asked to comment on this in the initial and the end of course questionnaires. The initial questionnaires were given to the students to complete a few months into the engineering phase and the end of course questionnaires were distributed at the end of the course just prior to the industrial attachment phase. Their responses are represented graphically in Table 5.11.

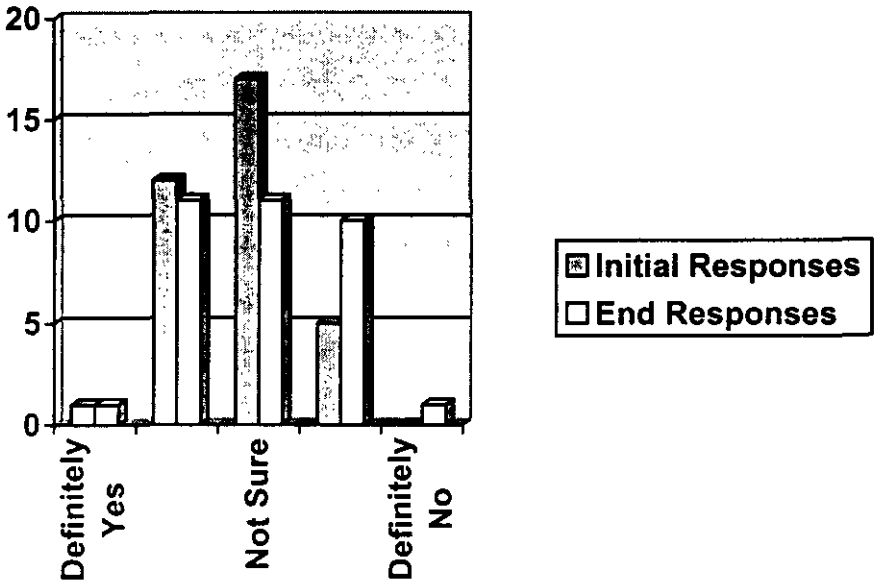


TABLE 5.11. Student response to question 13 from the initial and end of course questionnaire.

The graphical representation of the student responses to question 13 indicates an interesting trend. Between completing the initial and end of course questionnaires the students had completed the allied and engineering phases of the courses. The students when completing the end of course questionnaire had carried out, and been assessed on, the design projects and the integrated final design project. The earlier evidence showed that the students thought the projects were the assessment method that would predict future occupational performance. However, having completed the projects, some of the students revised their opinion of the course assessment with a clear shift towards the view that their potential to perform well in their future roles had not been assessed successfully. As the student questionnaires were returned anonymously the exact changes in individual responses cannot be identified but the downward shift is clearly evident. Further research carried out to identify any further changes in their views when the students were 6 months into their jobs as Clerks of Works would have been valuable but was beyond the scope of this study.

Employing Officers' and Serving Clerks of Works Perceptions

Although I was unable to carry out a survey of the perceptions of serving Clerk of Works engineers and their employing officers I was able to obtain access to a Review of the Clerks of Works employment and training carried out by the Training Design Team (TDT, 1999). I have chosen to include the an analysis of the Review as its findings highlight a number of issues and provide other evidence of the sort that Messick (1989) sees as supporting the claim for validity of the assessment. The results of the assessment are interpreted by the lecturers as inferring engineering competence on those that pass and they are used to post the newly qualified engineers into specific military jobs. As stated earlier, the lecturers' considered the courses produced individuals competent to carry out military engineering tasks for the Field Army but what other evidence is there that supports their view?

The Review stated that the Clerk of Works engineers were highly regarded for their high technical standards by all employing officers whether these officers were Army, RAF or civilian. This was reflected in the following quotes taken from the Review (TDT, 1999, p. 12) and regarded by the authors as typical comments from employing officers:

> "They [*the Clerk of Works engineers*] provide a unique capability which, if you [*the British Army*] loose it, you will not be able to regenerate it easily."
>
> (German Army Officer in Bosnia)

> "Clerks of Works are worth their weight in gold."
>
> (British Army Officer)

> "My Clerks of Works worked absolutely flat out in Bosnia... I could have used another 10... the hardest workers in the Regiment."
>
> (Head of G4 Estates, Civil Servant)

"Clerks of Works are essential to me and essential to the Army." (Deputy Garrison Commander, RAF Officer)

The Review indicated that experienced Clerk of Works engineers were considered to have played a key role in the success of the Corps of Royal Engineers in recent years, notably in Northern Ireland and Bosnia. In the case of Bosnia the Clerk of Works engineers earned an enviably high reputation in the eyes of other NATO nations. They were seen by employing officers as the basic building blocks of the specialist component of the Corps, providing the necessary supervisory and managerial engineering expertise without which the regiments could not carry out their military tasks. This was evidence that supported the lecturers' view that the Army did get military engineers who were considered competent in their roles as Clerks of Works. However the Review presented evidence to suggest that the newly graduated Clerk of Works engineer did not arrive in the Field Army ready made for such accolades. Before I discuss this evidence I will deal with the issue discussed earlier in this chapter, how the requirements for the BTEC HND and the Army Training Objectives were dealt with.

The Review reported that the retention of the BTEC HND accreditation was essential as the award was seen as a recruiting incentive, a vital prerequisite for subsequent qualification as an Incorporated Engineer and as a method of ensuring the training delivered reflected current civilian best practice. The Review also acknowledged that there was a difficulty with the current Army training objectives for the courses that needed to be addressed. It was stated that the current training objectives were neither valid nor auditable. The Review recommended that the courses must in future have bona fide training objectives in full accordance with the Systems Approach to Training methodology. As stated earlier, the process of producing new training objectives is underway but the lecturers remain sceptical about the outcome. Due to the recognized inadequacy of the existing training objectives the Review stated that the individual lecturers had considerable latitude to change aspects of the curriculum without formally making any concomitant amendments to the existing course training objectives. It was seen as essential that the majority of the Clerk of Works course lecturers

should therefore be military to *ensure that the appropriate interpretation of civilian standards and their military application is taught* (TDT, 1999, p. 3). The Review indicated that the practical experience of the lecturers was viewed by the Army, the employing officers, and serving Clerks of Works as being critical to the process of course design and assessment. The implication is that support was being expressed for the system described in the earlier sections of this chapter, that the BTEC HND set and validated the standard for the academic content and the military lecturers set and measured the standard for occupational performance. But does this symbiosis by itself produce the Clerk of Works described at the beginning of the section?

The lecturers' claim was that they produced what the Army wanted but that the newly graduated Clerk of Works engineer still lacked the experience needed to develop their personal knowledge. The Review stated that there was an almost universally positive view of the Clerk of Works engineers' ability to produce a flawless technical report. However, it was reported that newly graduated military engineers had a tendency to perceive that this was the only acceptable approach. Although the employing officers were very impressed they expressed a

> discernible vein of frustration about the preference of some Clerks of Works for the most technically perfect and elaborate engineering solution no matter how minor the problem. (TDT, 1999, p. 28)

According to the Review, employing officers stated that solutions produced were sometimes too expensive and utterly impractical to build within the limits of time, money, labour, and the materials that were actually available. The Review stated there was evidence that the employing officers said that some Clerk of Works engineers produced reports that went into immense detail and were excessively lengthy. The reason the Review cites for this can be summed up by the following quote.

> Clerks of Works training at the RSME places great emphasis on technical excellence and the highest possible technical standards. After almost two years of being imbued with this ethos it is not surprising that the newly graduated Clerk of Works has a tendency to perceive that this is the only acceptable approach.
>
> (TDT, 1999, p.29)

In contrast to the comments about the newly graduated Clerks of Works, it was stated that the more experienced Clerk of Works engineers exhibited a more pragmatic and flexible attitude towards engineering solutions and writing of the accompanying reports. They had taken a *reality pill* that *brought them down to earth in design, report writing and implementation matters* (TDT, 1999, p. 30). So the Clerk of Works engineers were not worth their weight in gold until they had developed personal knowledge gained from experience. The evidence presented in the Review suggests that graduating Clerk of Works engineers had the technical ability to produce a flawless solution but lacked the experience to see what the real time solution should be. A view that supports the earlier evidence, the importance Vincenti (1993) gives to practical experience in the development of engineers, and the distinction between experts and novices.

Summary of the Chapter

In this chapter I have presented and analysed data that has enabled the collection of evidence from which conclusions can be drawn about the validity of the assessment. The chapter began by looking at the way the Army Training Organisation, BTEC and the lecturers dealt with the concepts of validity moving on to look at the Army and BTEC systems for defining the construct. The main part of this chapter focused on the assessment process. The assessment process was broken down into three sections, the course assessment programmes, the assessment methods and how the lecturers carried out the assessment. The final section presented the data from other sources, i.e. the students, serving Clerk of Works engineers

and their employing officers. I will now draw on the evidence presented in this chapter to present my conclusions.

CHAPTER 6

Conclusions

The conclusions of my study are divided into three sections. The first section presents my conclusions about the validity of the assessment. The second section presents what I consider to be the implications of my findings to those involved with occupational training and education at the RSME and at other educational institutions. In the third section I will discuss the implications that have arisen from the process of conducting the study. This section will be of relevance to educational researchers concerned in particular with exploring issues of assessment and validity.

Conclusions from the Data and its Analysis

The aim of my study was to test my hypothesis that the assessment of the Clerk of Works students was valid and to establish the nature of that validity. I set out to establish the source of validity, whether the assessment was, in some form, valid due to the explicit intention of those responsible for the assessment process or due to the nature of the assessment process inherent in the organisation. The Army Training Organisation and BTEC defined a valid assessment as one that measures what it is supposed to measure, a simple statement that would appear to be easy to understand and easy to meet. Guidance from the Army Training Organisation and BTEC on how to ensure an assessment was valid, either in terms of designing a new assessment or evaluating an existing assessment, was limited. Of the types of validity discussed in Chapter 3, only face validity was identified as a particular type. There was no reference by the Army Training Organisation or BTEC to the three main types of validity defined by Guion (1980) and Wood (1991) as content, construct and criterion-related validity. The Army and BTEC focused on the need for sampling of the domain by use of a number and range of different assessment methods. In these terms a valid assessment measures, through adequate sampling, what it is supposed to measure. If the *what* is a well defined construct then it would be reasonable to conclude that validity was being conceptualized as construct validity. If

what is supposed to be measured is the construct and this is clearly defined and measured then the assessment will be valid and the results can be used to place graduates into the Field Army as Clerk of Works engineers. The problem lies in defining the construct and hence domain specification, a problem area that was shown on page 31 to have been widely recognized by those concerned with competence-based systems.

The Army system of defining the construct based on a job analysis and the development of training objectives appears on paper to meet the requirements of a competence-based system perfectly. The Army focus was placed on the key purposes and functions of the occupation, a key requirement highlighted by Eraut (1994) and Wolf (1995, 1996). If the construct was clearly defined and assessed by effective simulation and sampling then the evidence from the studies critiqued by Wolf (1995) suggests that the Army system, supported by the BTEC system, would provide a valid form of assessment of occupational performance. However the evidence presented from the lecturers, the TDT Review (1999) and the Army Training Organisation themselves, through their publications, showed that the construct military engineering competence could not be clearly defined using the Army Systems Approach to Training process. The process could not specify the domains clearly and unambiguously. The domains identified, cognitive and affective, matched those from Imrie's (1995) discussion of engineering taxonomies, however the Army Training Organisation stated that they believed it was only possible to be precise both in statements of required behaviour and in the measurement of achievement for simpler levels of knowledge and procedure. It was evident from the data presented that between 1976 and 1998, apart from an ill-fated attempt in 1989, no serious effort had been made to produce a definitive set of training objectives. The accreditation of the course with BTEC supplied a generic model around which the lecturers could base the content of the course but this could not, and indeed did not lead directly to a definition of the construct military engineering competence. The responsibility was passed to the lecturers as subject matter experts and experienced Clerk of Works engineers to define the construct based on their own internalized model and they did not refer to the guidance from either the Army Training

Organisation or BTEC. This leads to the conclusion that neither the Army Training Organisation nor BTEC can be considered to be an explicit source of the validity of the assessment.

The lecturers considered the main aim of the course was to produce a competent military engineer able to do the job of a Clerk of Works. The award of the BTEC HND was seen as providing a means of comparison with other military and civilian courses and as an incentive to the students to come on the course as they would gain a recognized civilian qualification. The amalgamation of the occupational requirement and the generic academic requirement resulted in an emphasis on the academic rather than performance measures. The evidence from the analysis of the assessment process indicated that the lecturers had not clearly defined the construct military engineering competence in terms of specified domains, performance criteria, and the standards to be met. The lecturers did not approach the assessment process with the explicit aim of ensuring validity was designed into the process. The lecturers therefore cannot be considered to be an explicit source of validity and hence the conclusion drawn is that it is the inherent characteristics of the assessment process that provides the source of the validity of the assessment.

It is at this point that the theoretical criteria, developed from Messick's (1992) set of questions, against which I set out to evaluate the validity of the assessment can be addressed. The first four of the theoretical criteria, shown below, focused on the assessment itself, and dealt with the evidential basis of validity in terms of Messick's (1989) four facets of validity that make up the unitary concept.

- The construct is represented in the assessment in sufficient breadth and depth.
- Important elements of the construct have not been omitted from the assessment.
- The assessment does not introduce sources of invalidity or irrelevant variance that bias the scores or judgements.

- The way the assessment is marked reflects the manner in which domain processes combine to produce effects and is consistent with the structure of the domain about which inferences are to be drawn or predictions made.

The evidence indicated that the construct, military engineering competence, was not well represented in the assessment. There was sufficient evidence to suggest the subject domains, e.g. mathematics, physics, heating, electrical transmission and distribution, were well represented both in the written examinations and in the design projects. In this respect, the first four of Vincenti's (1993) categories of engineering knowledge discussed in Chapter 3, i.e. fundamental design concepts, criteria and specifications, theoretical tools, and quantitative data, were well represented in the assessment. The responses from the students indicated that they felt all the assessment methods assessed what was taught in class and due to the use of textbooks and BTEC unit specifications I feel that what was taught represented the academic curriculum content, the subject domains. There was some evidence that the lecturers were attempting to assess cognitive skills and behavioural outcomes, the cognitive and affective domains, but not in a formal manner using clear, unambiguous outcomes or criteria against which performance was assessed and recorded. The design projects represented the simulated work based tasks and the way that they were marked was not consistent with the inference that the projects were assessing the construct military engineering competence. The absence of any clear performance criteria and standards against which to measure the criteria meant that the only way that the marking system would be a consistent measure of the construct would be through the use of the lecturers' judgement. There was some evidence of compensation occurring but there was insufficient evidence to conclude that bias or compensation of the marks awarded for the project was taking place. The lack of Army training objectives meant that the assessment could not be structured around the results of the job analysis system. The BTEC unit specifications provided the lecturers with a framework on which to base their teaching and assessment of learning, there was a requirement for accreditation of the courses, and visits from the BTEC moderator involved reviewing the assessment of the students. The

use of written examinations and project marking guides that focused on technical design solutions made marking and aggregation an easier formal process to manage. It was very difficult to find evidence of construct irrelevant variance in the assessment, as in many cases I was not a subject matter expert. The evidence from the students suggests it is possible there was some degree of irrelevant variance present in the examinations and projects. However I have no supporting evidence of this and I feel any conclusions I may draw with regard to irrelevant variance could not be substantiated from the evidence presented.

There was evidence to conclude that the assessment had content validity, the assessment measured what was taught but what was taught was essentially from the subject domain and linked to the BTEC unit specification and the textbooks. The problem with drawing conclusions based on the evidence presented is that at no time was data collected from the teaching environment and therefore no empirical evidence has been presented in support of this conclusions. The content of the project marking guides indicated that the students were given marks for technical content and not how they carried out the design individually and as teams, how they conceptualized and how they communicated their design solutions. The assessment methods used were appropriate as a measure of student learning, the use of laboratory assignments, written examinations, design projects and the industrial attachment showed that a range of methods were used. The methods were relevant to the assessment of military engineering competence and presented no problems in terms of their utility. However there was evidence that showed the way the design projects were marked did not meet the last of the four criteria listed earlier. Although the lecturers described elements of the assessment process where they assessed cognitive skills and behavioural outcomes, and this was supported by student data, the documentary evidence showed few, if any, marks were recorded against these elements. The lecturers could be seen as an implicit source of validity due to the use of their professional judgement but the processes that would have ensured explicit consideration of the construct and consensus among the lecturers were not sufficient. Hence the implicit element that rested on the use of the lecturers' professional judgement could not be considered

adequate to ensure validity of the assessment. The findings from my study lead to the conclusion that the construct was clearly defined in terms of academic curriculum content but not in terms of cognitive skills and behavioural outcomes. The construct military engineering competence was not being represented adequately in the assessment and any claim made for construct validity would be unsubstantiated.

The next issue is that of the interpretation and use of the results of the assessment and the associated consequences. The last four theoretical criteria were concerned with interpretation and use of results and the consequential validity of the assessment.

- The results of the assessment can be interpreted to mean students that pass have demonstrated the required level of performance to be considered competent military engineers.
- The results of the assessment can be used to place graduates of the courses into jobs in the Field Army.
- A consequence of interpretation and use of the results of the assessment is that graduates go on to perform successfully in their occupational role.
- A consequence of interpretation and use of the results of the assessment is that there is a positive impact on the teaching and learning process.

The results of the assessment could not be interpreted to indicate that graduates had demonstrated the required level of performance to be considered competent military engineers. There was strong evidence to suggest that the graduates could be classified as novices lacking in the personal knowledge required to become expert Clerk of Works. The lecturers and the students recognized the lack of experience and the fact that it limited the scope and quality, using Eraut's (1994) terminology, of what they were able to do. The evidence from the TDT Review (1999) reinforces the point that before they become experienced the newly qualified Clerk of Works rely on their ability to produce a technically correct design solution in the way they had been taught to do through the use of the design projects. So could the results be used to place graduates in jobs as Clerk of Works in the Field Army? The conclusion is yes, the results could be used in this way

but with reservations. The evidence from the TDT Review (1999) suggests that the expectation of the employing officers was that the graduates would arrive competent to do the job and this included having developed sufficient personal knowledge. In view of the evidence, this expectation was too high to be achieved. The claim made by the lecturers that they produced what the Army wanted was supported by the employing officers' view of the experienced Clerk of Works. The implication was that a consequence of the interpretation and use of the results was that graduates go on to perform successfully and therefore there must be a strong correlation between the result of the assessment and future occupational performance. The result of the assessment is derived from academic measures and as Wolf (1995) stated from her critique of the relevant studies the correlation between academic measures and occupational performance is low. There was agreement between the lecturers and the students that the results of the assessment, particularly in the foundation phase, were a predictor of the students' ability to pass the course. Their views were supported by the study of Williams and Boreham (1971). From the available evidence, my conclusion is that the majority of graduates, after experience in the job, do go on to perform successfully as Clerk of Works engineers. However, the successful future performance of the graduates is not necessarily as a result of the assessment and I would conclude that the results of the assessment should not be interpreted and used as a predictor of future performance. Further research, discussed in the next chapter, is required in this area but was beyond the scope of my study.

The view was expressed in the theoretical discussion that systematic validity of the assessment as described by Frederiksen and Collins (1989) would be present if there was a positive impact on teaching and learning as a result of the interpretation and use of the results of the assessment. The students were learning to be military engineers by doing the projects and through the interaction with the lecturers, formally and informally, and there was evidence of the directness of cognitive assessment and judgement in assigning a score. Seeking evidence for systematic validity required looking at teaching, learning and assessment in much greater depth than took place in my study. Hence I feel I have insufficient evidence to draw any

conclusion about the systematic validity of the assessment.

Having evaluated the validity of the assessment against the theoretical criteria what conclusion can be drawn about the validity in terms of the unitary concept? The construct being assessed was military engineering competence and the claim being made for the results of the assessment was that the results could be interpreted and used to place graduates into jobs as Clerk of Works engineers. The consequence of the use and interpretation was that the graduates required a period of experience before they were considered to be good Clerk of Works. In light of the evidence presented, the theoretical criteria against which the validity of the assessment was being measured were not satisfied sufficiently to conclude the assessment was valid in terms of the unitary concept.

The Implications of the Findings from My Study

There are implications from the findings of my study that are significant to the Army, BTEC and the lecturers involved in assessment of the students at the RSME and to educators and trainers working in other educational institutions. The implications are centred on four areas, defining the construct, the use of subject matter experts in the assessment process, accreditation of occupational courses for civilian qualifications, and the claims made for the assessment in relation to future occupational performance.

The findings from my study emphasized the problems associated with defining and assessing competence as a construct. Defining the construct military engineering competence in terms of defined performance criteria that were meaningful and assessable did not occur. There were no clear criteria to measure cognitive skills or behavioural outcomes. The construct was not being specified clearly and unambiguously and the way around this problem was seen to be to employ experienced Clerk of Works engineers, experts, as lecturers. The military experts, the people with experience of doing the job, were given the task of training and assessing the students. The purpose of employing these *experts* in the role of the deliverers of the

training can be seen as twofold, to define the construct and to assess the construct. A number of assumptions are being made here.

- The experts can define the construct explicitly.
- They can teach to the construct they define.
- They can design an assessment that will be a valid assessment of that construct.
- They can define the performance standards against which to measure the performance criteria developed from the construct.
- They will not bias or compensate student results based on their own individual internalized, holistic models of the construct.

The findings from my study indicated that the expert engineers did not meet the requirements of any of the above statements. The implication from my findings is that all the above are difficult to do when dealing with constructs such as military engineering competence. If subject matter experts are to do these things successfully then there must be an appreciation that they are experts in their field, not experts in competence-based assessment. Therefore I fully support the views of Burchell *et al*, (1999), Dunbar *et* al, (1991), Eraut, (1994), Linn, (1993), and Kane, (1994) that the training of these experts, who use professional judgement to define the construct and design assessment, has a critical part to play. The training does not need to lead to an understanding of the theoretical discussion on concepts of validity presented in my study. If the training results in clearly specified domains that are represented adequately in the assessment then the concept of a valid assessment measuring what it is supposed to measure would suffice. The main point is that appointing subject matter experts is not enough, they must be given training and guidance if they are to come close to achieving what is expected of them i.e. a valid assessment of the construct.

There are implications that have arisen from the accreditation of the course with BTEC. On the positive side accreditation introduces a means of externally moderating the course and this would bring with it comparability and, perhaps, reliability of the assessment. The findings from my study suggest that the amalgamation of an occupational engineering course with a

generic engineering course had a positive and negative impact. The BTEC accreditation and use of the unit specifications supplied a structure to the course and brought with it a certain academic rigour. In doing so, the positive impact was that the reliability and comparability of the results of the assessment was likely to be enhanced. The negative impact was that the course content was structured around the BTEC units and although the BTEC moderator and the lecturers suggested design projects played a major role in the assessment, the assessment relied heavily on academic measures. The implication of the BTEC involvement was that the assessment moved away from rather than towards assessing performance criteria that related to the construct, military engineering competence. Wolf's (1995) critique of studies concerned with the correlation between the results of the assessment and occupational performance would support the view that the incorporation of the BTEC HND resulted in an assessment process that reduced the likelihood of the results being a good predictor of future occupational performance. There is justification for accrediting occupational courses with qualifications like the BTEC HND. However care must be taken to ensure the effect of accreditation is not to move away from the purpose of the assessment in terms of the how the results of the assessment are to be interpreted and used.

There is a very clear implication that has emerged from my findings and it is concerned with what claims can be made for an assessment of occupational competence carried out at an educational establishment. There must be a match between what the results of assessment really mean and what they are interpreted to mean. Competence should be viewed in terms of scope and quality, as defined by Eraut (1994) and discussed on page 30, and the notion that competence should be based on a model of progression and lifelong professional learning. There was overwhelming support for Vincenti's (1993) view that the personal knowledge required of an engineer cannot be developed in a classroom setting. The industrial attachment at the end of the course was an acknowledgement of this problem and supported Wolf's (1996) view that assessment could be substituted for periods of experience. The implication from my findings is that the inclusion of a period of work experience should be considered an essential aim for all occupational

courses of a similar nature to the Clerk of Works courses.

The Implications from Conducting My Study

There are two related implications that arose out of conducting my study that I consider are relevant to those educational researchers conducting studies into assessment and its validity. The first implication was that although the initial aim was to conduct a qualitative study employing an interpretive approach it became evident that the study would in fact rely on both qualitative and quantitative methodology. Although this issue has been discussed in Chapter 4 and the rationale explained in terms of analysing the data and presenting it within a military culture, there is an implication that arises out of my study. It would have been very difficult to conduct a study into the validity of assessment in terms of the unitary concept due to the number and complexity of the issues involved without employing both qualitative and quantitative methods.

In the introduction to Chapter 3 I stated that validity of assessment was widely acknowledged as a very complex and problematic concept. The development of the theoretical criteria I used in my study that were developed from Messick's (1992) set of questions and designed to evaluate the validity of the assessment in terms of the unitary concept further indicated the complexity of the issues. The discussion of the works of Frederikson and Collins (1989), Linn, Baker and Dunbar (1991), Hickey, Wolfe and Kindfield (1999), and Shepard (1993) highlighted this still further. Carrying out my study has confirmed for me the relevance of Gipps' (1995) view that using Messick's model leads to the need to address validation requirements of extensive proportions. Although I tried to collect data that related to all the theoretical criteria I had developed, generating both qualitative and quantitative data in the process, it became evident that individual studies would be needed to address adequately some of the criterion. Examples of this would be the need to study the teaching and learning process to address the issues of construct representation and systematic validity, and to study longitudinal occupational success to address issues of consequential validity. The implication from my study to

other educational researchers is that the validity of assessment is indeed a complex and problematic area. Careful thought should be given to the aim of the research, the form the study will take, and the breadth and depth of any exploration into assessment and its validity.

CHAPTER 7

An Evaluation of My Study

Introduction

In this, the final chapter, I will reflect on and evaluate my study in terms of the process of conducting it, the contribution to the theory and practice of education, and the identification of areas for future research. In evaluating the process I will look at how I could have approached areas of the study differently, how some opportunities were missed and how others could have been explored further. Then I will discuss how my findings have, or will be, presented to the RSME and how I intend to circulate it to a wider audience. Finally, I will discuss possible areas for future research.

Evaluation of the Process

The first lesson I have learned is that to conduct quality research and produce work of value to the community to which it belongs, the researcher must develop or already possess three things. The first is a thorough grasp of the situation within which the research is to be conducted. The second is a sound theoretical knowledge of the subject areas being researched. The third, arguably the starting point from which the research begins, is an idea of the hypotheses, and/or the research questions, to be investigated.

My own starting point for the journey that lead to the production of this dissertation was that I wanted to explore the issues concerned with the validity of assessment. My research began with the simple question in what way was engineering assessment valid. This was a question that was extremely relevant to my work at the RSME as I trained engineers from HND up to MSc level. It was of interest to me personally as an engineer, and in attempting to answer it I felt I would be producing something of value to those concerned with the training and assessment of engineers, both at the RSME and in general.

As I researched the subject area I very quickly became aware that I had chosen to address a question to which the answer was in fact far from simple. The first problem I had was the inability to find any studies that dealt with the issues in a similar manner to the study I had begun to carry out. This was both in terms of topic area, the way I was approaching the issues and the fact that I was dealing with a unique situation, i.e. the military training environment. It was at this point I realized the importance of developing a strategy for dealing with the collection and assimilation of the current theory and practice contained in accessible literary sources. At first I thought I needed only to concentrate on the literature concerned with validity and assessment however it soon became evident that I needed to look at issues of competence-based assessment and views on the nature of engineering competence as a construct. One point I reflected on many times during my research was that all the issues I was exploring were extremely difficult and complex. It became clear that it was very easy to develop an extensive reading list, especially as each area branched out, and it became a necessity to establish the relevance of the literature to the study. I read considerably more literature than I have discussed in Chapter 3 and Chapter 4 as this was necessary in order to establish a greater understanding of the literature used and its relation to my study. It was also very clear that the literature review had to continue throughout the research so that I did not miss any recently published works relevant to my own study and therefore time had to be allowed for this in my research programme. As my study progressed I found I became more efficient at the analytical process, extracting the key issues and arguments from a literary text. I believe this was helped by the fact that I was becoming more knowledgeable in the subject areas and hence more aware of exactly what issues I was trying to explore.

As well as gathering the literature on the subject matter I was also aware of the need to look at the methodological issues in a similar manner. I found this equally challenging as I became more aware of the quantitative versus qualitative debate and the issues involved in choosing a particular methodology. I feel that I now have a far greater understanding of these issues, as outlined in Chapter 4, and an insight into the world of research

that I did not have before I began my study. What also came out of this was the recognition of the need to choose the right methods for the collection of the data and its analysis. Choosing the methods did not seem to be too difficult as I felt my study would involve documents, interviews and questionnaires and that these were rather obvious choices. I found I did not have any difficulty identifying the relevant documents nor analysing their content. This was almost certainly due to the learning and experience I gained from carrying out my literature review.

I felt that the data collected from the interviews improved as I developed a greater feel for what I needed to ask. After the first three interviews they became considerably longer, more relaxed and more focused on the issues I wanted to explore. I became more aware of how to follow up a line of questioning and in recognizing the need to probe more deeply to establish exactly what the interviewees meant when using particular terms and phrases.

Gathering student data by the use of questionnaires proved to be the most enlightening of the three methods. My first questionnaire was issued during the initial study stage of my research because I wanted to test it out and I had to try to get as much data as possible from as many courses as possible. I thought I had put together a reasonably good set of questions however it became clear to me that it was in fact poorly constructed. The questions were too open, one question was mis-interpreted and another gave the students the opportunity to give positives only rather than a balanced view. In terms of the questions being too open I felt that I was asking for too much in the way of descriptive answers from the students. The outcome of this was twofold, first it made it harder and more time consuming for the students to fill in, and second it made it more difficult for me to assimilate the data produced. This all resulted in spending considerable time and effort in developing the two questionnaires included in the appendices.

As is probably true for most researchers producing a work such as this, I am still a little dissatisfied with what I have achieved. Although I focused on assessment and collected a great deal of data I missed the opportunity to

look in detail at the attachment phase of the courses. I did not at first recognize the need and also there were considerable time and travel implications. Site visits to observe the students at work, to speak to them, their civilian peers and their industrial manager would have been invaluable to me in the discussion on performance and hence competence. I would have been able to get the views of those working in industry on how well the students performed on a range of tasks they dealt with during the attachment.

Although I looked at the Army training objectives and the BTEC unit specifications I did not look at what the lecturers taught in the classroom. This was because I had decided that their view of the construct would be better expressed from how and what they assessed rather than what they taught. I still maintain this view, however observing what the lecturers presented in class and the interactions between them and the students might have produced useful data. It would have provided an insight into the consequences of the assessment process in terms of the teaching and learning process. I also feel I should have observed the students in the process of carrying out their projects. I think this is more an area I could have explored further rather than a missed opportunity. I did observe the process of assessment of the student projects i.e. the presentations and meetings, because I was looking at this as part of the assessment process.

There were a number of times when I referred back to my initial study report submitted as part of the EdD assessment process and used it as the guiding reference for my research. However much has been developed since then and when I compare the content and style to this, the final dissertation, I am quite astonished by the changes I can see that have occurred. Obviously the content in all areas has been developed considerably but it is the way the content has changed and the way it is now presented that show me how much I have developed my own thoughts, ideas and arguments. I now recognize how important it is to be able to present them logically and coherently. I realize that this research has also been about the developing my ability to conceptualize. Even now after reading the completed dissertation there were still areas where I felt I could have structured it

differently or added some additional data, analysis, or thoughts that might have improved it still further. So am I a better researcher now than when I began the study? The answer has to be a definite yes.

Evaluation of the Contribution to the Theory and Practice of Education

All EdD students from my cohort attended a weekend conference and were required to deliver a presentation on their research. At the end of my presentation I was asked in what way my study of Army training contributed to the wider community of educational theory and practice. My answer to this was that although the context of my study was military, most of the issues contained within it applied equally as well to civilian educational institutions. In fact, I believe that the culture within which my study is embedded is one that has many rich veins of interest to educators and my study is an attempt to deal with just one. I have explored a number of complex issues in terms of the validity of assessing engineering competence and attempted to deal with them in such a way as to make the ideas more accessible and more easily understood. My conclusions in the previous chapter highlight the implications of my study to the theory and practice of education and my study has already resulted in a number of positive moves forward at the RSME. These are:

- A job analysis and review of the training objectives which now involves the lecturers in the process.
- The assessment processes for both the electrical and the mechanical courses are now being reviewed.
- Meetings between the lecturers are taking place on a regular basis to discuss assessment issues.
- As a result of the interview process a number of lecturers have reviewed their own input into the assessment process and also have acknowledged that the validity of assessment needs to be given proper attention.
- In my view, as an advocate of reflective practice, the influence my study has had on all the lecturers has been very positive and it has resulted in a

teaching and learning environment where reflective practice now plays an important part. The lecturers have, themselves, become reflective practitioners.

In terms of publication and dissemination of the content of this dissertation there are two areas. The first is publication to the military audience. Much of my study has already been discussed informally with the lecturers and the management at the RSME and a formal presentation will take place towards the end of the year. A copy of my dissertation will be placed in their library as a reference document. I also aim to publish an article in both the Army Doctrine and Training News Journal and the Defence Management Journal. In terms of accessing a wider public audience I will be looking to publish in appropriate journals e.g. the European Journal of Engineering Education, at a later date.

Identification of Future Research

For me, this study was an exploration into the world of the curriculum, assessment and its validity. What I uncovered has certainly lead to a greater understanding of the issues on my part and highlighted a number of future areas for research. I have no doubt that the nature of military training carried out at the RSME presents many opportunities for further research in areas that would add value to the theory and practice of education. All the training carried out is occupationally focused, involves the job analysis process, the award of a recognized qualification and ranges from basic trade training up to MSc level. Those who pass the courses take up jobs within the Field Army and most will return to the RSME for further training at some point in their careers. This all points to a number of areas for further research and these are outlined below.

- The culture of the organisation and the nature of the training make it an ideal environment where longitudinal studies of occupational success could be carried out successfully. The majority, if not all, of the students from my study will still be serving as Clerk of Works engineers for at

least four years. Researching their development would be one area for future consideration.

- The trade training courses, where students return over a number of years for class III, II and I trade training would provide a rich source of data in a number of areas, for example studies in assessment, learning, competence and occupational performance.
- Participation in the community of practice and the transfer of skills is another area. A follow up study that focuses on how the students transfer learning from the training establishment to their job in the Army would be informative and interesting to conduct. This was hinted at in the TDT Review but I believe there is a lot more that could be gained from a focused study of the changes the graduates go through and how they transfer learning to the working environment.
- There is an internal selection process that is conducted for the trade courses and for the Clerks of Works courses and this was not researched during my study.
- My study has shown that there is great difficulty in defining the construct and assessing it in terms of performance criteria. I established that informal assessment was taking place of key elements of the construct. I was not concerned with looking at how, for example, the affective domain could be defined and formally assessed but I now think this is an issue worthy of further exploration.
- The teaching and learning process, what went on in the classroom, was not studied in this dissertation and would certainly be an area for future research. Especially as it could add to the notion of assessment being a tool of the curriculum.
- The uniqueness of my study was that it was conducted at a military training establishment. It was a single case study. Research into the same issues but at other, similar establishments both military and civilian would be another step forward.

I believe there would be much to be gained from further research into any of the areas discussed above and that my own research serves as an excellent starting point for any such journey.

References

Atkinson, P., Delamont, S. and Hammersley, M. (1993) 'Qualitative research traditions' in Hammersley M. (ed.) *Educational Research - current issues*. London, Paul Chapman Publishing.

Ashcroft, K. and Palacio, D. (1996) *Researching into Assessment and Evaluation in Colleges and Universities.* London, Kogan Page.

Asher, J.J. and Sciarrino, J.A (1974) Realistic work sample tests: a review. *Personnel Psychology* **27**, 519-33.

Army Training Organisation (1988) *The Validation of Training - Pamphlet No. 4.* Upavon, Army School of Training Support.

Army Training Organisation (1993) *The Writing of Training Objectives - Pamphlet No. 3.* Upavon, Army School of Training Support.

Army Training Organisation (1995) *The Use of Testing in Training - Pamphlet No. 5.* Upavon, Army School of Training Support.

Baker, E. (1992) *The Role of Domain Specifications in Improving the Technical Quality of Performance Assessment.* CRESST, UCLA.

Benett, Y., Lee, B. and Jackson, J. (1989) 'The development of placement and assessment procedures in supervised work experience', *Project Report No 31.* London, CNAA.

Benett, Y. (1993) 'The Validity and Reliability of Assessment and Self-assessment of Work Based Learning', *Assessment and Evaluation in HE* **18**(2), 83-94.

Bird, M. and Hammersley, M. assisted by Gomm, R. & Woods, P. (1996) *E835 Educational Research in Action - Study Guide.* Milton Keynes, The Open University.

Boud, D. (1995) *Assessment and Learning: Contradictory or Complimentary?* in Knight, P. (ed.) *Assessment for Learning in Higher Education.* London, Kogan Page.

Brown, G., Bull, J., and Pendlebury, M. (1997) *Assessing Student Learning in Higher Education.* London, Routledge.

Brown, S. and Knight, P. (1994) *Assessing Learners in Higher Education.* London, Kogan Page.

Bryman, A. (1992) 'Quantitative and Qualitative Research; further reflections on their integration' in Brannen, J. (ed.) *Missing Methods.* Aldershot, Avesbury Press.

BTEC (1986a) *Assessment and Grading - General Guideline.* London, BTEC.

BTEC (1986b) *Course and Unit Design - General Guideline.* London, BTEC.

BTEC (1988) *Course Review and Education - General Guideline.* London, BTEC.

Bucciarelli, L. L. (1994) *Designing Engineers.* London, MIT Press.

Burchell, H., Higgs, T,and Murray, S. (1999) 'Assessment of Competence in Radiography Education', *Assessment and Evaluation in HE* **24**(3), 315-26.

CIBSE (1999) *Building Services Rule of Thumb.* London, Chartered Institution of Building Service Engineers.

Cotton, J. (1995) *The Theory of Assessment: An Introduction.* London, Kogan Page.

Court, A. W. (1998) 'Improving Creativity in Engineering Design Education', *European Journal of Engineering Education*, **23**(2), 141-54.

Cox, A. and Harper, H. (2000) *Planning Teaching & Assessing Learning - A Reader, School of Post-Compulsory Education & Training.* London, Greenwich University Press.

Cronbach, L. J. (1988) 'Five Perspectives on Validity Argument' in Weiner, H. & Brown, H. (eds.) *Test Validity.* Princeton, NJ, Erlbaum.

Cronbach, L. J. (1989) 'Construct Validation after Thirty Years' in Linn, R.E. (ed.) *Intelligence: Measurement, Theory and Public Policy.* Urbana, University of Illinois Press.

Cronbach, L. J. and Meehl, P. G. (1955) 'Construct Validity in Psychometric Tests', *Psychological Bulletin*, **52**, 281-302.

Dary Erwin, T. (1995) 'Attending to Assessment: A process for Faculty' in Knight, P. (ed.) *Assessment for Learning in Higher Education.* London, Kogan Page.

Denzin, N. K. (1989) *Interpretive Interaction.* London, Sage Publishing.

Dunbar, S.B., Koretz, D.M. and Hoover, H.D. (1991) 'Quality control in the development and use of performance assessments', *Applied Measurement in Education*, **4**(4), 289-304.

Easterby-Smith, M., Thorpe, R. and Lowe, A. (1994) 'Analysing Qualitative Data' in Bennett, N., Glatter, R. and Levacic, R. (eds) *Improving Educational Management through research and consultancy.* London, Paul Chapman Publishing.

Engineering Professors Conference (1989) *Quality in Engineering Education*, Engineering Professors' Conference (UK), Occasional Paper No. 5.

Eraut, M. (1994) *Developing Professional Knowledge and Competence.* London, Falmer Press.

Faulkner, D., Swann, J., Baker, S., Bird, and M., Carty, J. (1993) *Professional Development in Action - Methodology Handbook.* Milton Keynes, The Open University.

Feltham, R. (1988a) 'Validity of a police assessment centre: a 1-19 year follow-up', *Journal of Occupational Psychology*, **61**, 129-44.

Feltham, R. (1988b) 'Assessment centre decision-making: judgmental vs mechanical', *Journal of Occupational Psychology*, **61**, 237-41.

Fleming, N.D. (1999) 'Biases in Marking Students' Written Work: Quality?' in Brown, S. and Glasner, A. (eds.) *Assessment Matters in Higher Education - Choosing and Using Diverse Approaches.* Buckingham, Open University Press.

Fletcher, S. (1991) *NVQs Standards and Competence. A Practical Guide for Employers, Managers and Trainers.* London, Kogan Page.

Frederiksen, J. R. and Collins, A. (1989) 'A Systems Approach to Educational Testing', *Educational Researcher*, **18**(9), 27-32.

Futcher, G. (1989)'Measurement or Assessment: A Fundamental Dichotomy and its Educational Implications' in Murphy, P. and Moon, R. (eds.) (1989) *Developments in Learning and Assessment.* London, Hodder & Stoughton.

Gardner, K.E. and Williams, A.P.O. (1973) 'A twenty five year follow-up of an extended interview selection procedure in the Royal Navy', *Journal of Occupational Psychology*, **47**, 1-13 and 149-61.

Ghiselli, E.E. (1966) *The Validity of Occupational Aptitude Tests*. New York, John Wiley.

Gipps, C. (1995) *Beyond Testing: Towards a theory of educational assessment*. London, Falmer Press.

Glaser, R. (1963) 'Instructional technology and the measurement of learning outcomes: some questions', *American* Psychologist, **18**, 519-21.

Glaser, R. (1993) 'Education and Thinking: the Role of Knowledge' in McCormick, R., Murphy, M. and Harrison, M. (eds) *Teaching and Learning Technology*. Wokingham, Open University.

Goetz, J. and Le Compte, M. (1984) *Ethnography and Qualitative Design in Education Research*. Orlando, Academic Press.

Gonzi, A., Hager, P. and Athanasou, J. (1993) 'The Development of Competency Based Assessment Strategies for the Professionals', *National Office of Overseas Skills Recognition Research Paper No 8*. Canberra, Australian Government Publishing Service.

Guba, E. and Lincoln, Y. (1981) *Effective Evaluation: Improving the Usefulness of Evaluation Results through Responsive and Naturalistic Approaches*. San Francisco, Jossey-Bass.

Guba, E. and Lincoln, Y. (1982) 'Epistemological and methodological bases of naturalistic inquiry', *Educational Communication and Technology Journal*, **30**, 233-52.

Guion, R.M. (1980) 'Content Validity - the source of my discontent', *Applied Psychology*, **11**, 385-98.

Haertel, E. H. (1985) 'Construct Validity and Criterion-Referenced Testing', *Review of Educational Research*, **55**(1), 23-46.

Haertel, E. H. (1990) *From Expert Opinions to Reliable Scores: Psychometric for Judgement-based Teacher Assessment*. Boston, American Educational Research Association.

Haertel, E. H. (1991) 'New Forms of Teacher Assessment', *Review of Research in Education*, **17**, 3-29.

Hager, P., Gonczi, A. and Athanasou, J (1994) 'General Issues about Assessment of Competence', *Assessment & Evaluation in Higher Education*, **19**(1), 3-16.

Hager, P. and Butler, J. (1996) 'Two Models of Educational Assessment', *Assessment & Evaluation in Higher Education*, **21**(4), 367-78.

Hambleton, R.K, and Rogers, H.J. (1991) 'Advances in Criterion-Referenced Measurement' in Hambleton, R.K. and Zaal, J.N. (eds) *Advances in Educational and Psychological Testing*. Boston, Kluwer.

Hickey, D. T., Wolfe, E. W. and Kindfield, A. C. H. (1999) 'Assessing Learning in a Technology-Supported Genetics Environment: Evidential and Systematic Validity Issues', *Educational Assessment*, **6**(3), 155-96.

Hillier, Y. (1997) 'Competence Based Qualifications in Training, Development and Management', *Journal of Further and Higher Education*, **21**(1), 33-41.

Imrie, B. W. (1995) 'Assessments for Learning: quality and taxonomies', *Assessment & Evaluation in Higher Education*, **20**(2), 175-189.

Jessup, G. (1991) *Outcomes. NVQs and the Emerging Model of Education and Training*. London, Falmer Press.

Kane, M. (1994) 'Validating thePerformance Standards Associated with Passing Scores', *Review of Educational Research* **64**(3), 425-61.

Kvale, S. (1996) 'Examinations reexamined: Certification of students or certification of knowledge?' in Chaiklin, S. and Lave, J. (eds) *Understanding Practice: perspectives on activity and context.* Cambridge, Cambridge University Press.

Linn, R. (1993) 'Educational Assessment: Expanded expectations and challenges', *Educational Evaluation and Policy Analysis*, **15**(1), 17-28.
Linn, R., Baker, E. and Dunbar, S. (1991) 'Complex, Performance-Based Assessment: Expectations and Validation Criteria', *Educational Researcher*, **28**(8), 15-21.

Linn, R. and Baker, E. (1996) 'Can Performance-based Student Assessment be Psychometrically Sound?' in Boykoff Baron, J. and Palmer Wolf, D. (eds.) *Performance-based Student Assessment: Challenges and Possibilities* Chicago, 95th Yearbook for the National Society for the Study of Education.

McCormick, R. (1997) 'Conceptual and Procedural Knowledge', *International Journal of Technology and Design Education* **7**, 141-59.

McMahon, F. and Carter, E. (1990) *The Great Training Robbery.* London, Falmer Press.

Messick, S. (1975) 'The standard problem: Meaning and values in measurement and evaluation', *American Psychologist*, **30**(10), 955-66.

Messick, S. (1980) 'Test Validity and the Ethics of Assessment', *American Psychologist*, **35**(11), 1012-27.

Messick, S. (1984) 'The psychology of educational measurement', *Journal of Educational Measurement*, **21**, 215-38.

Messick, S. (1989) 'Meaning and Values in Test Validation: The Science and Ethics of Assessment', *Educational Researcher*, **18**(2), 5-11.

Messick, S. (1992) *The Interplay of Evidence and Consequences in the Validation of Performance Assessments*, Research Report ETS, July.

Miles, M. B. and Huberman, A. M. (1984) *Qualitative data analysis: a source book of new methods*. London, Sage.

Moss, P. A. (1992) 'Shifting conceptions of validity in educational measurement: Implications for performance assessment', *Review of Educational Research*, **62**(3), 229-58.

Nuttall, D. (1989) *The Validity of Assessments* in Murphy, P. and Moon, R. (eds.) *Developments in Learning and Assessment*. London, Hodder & Stoughton.

The Open University (1994) E824 *Study Guide - Educational Research Methods*. Milton Keynes, The Open University.

Polanyi, M. (1962) *Personal Knowledge*. Chicago Press.

Popham, W.J. (1978) *Criterion-Referenced Measurement*. Englewood Cliffs NJ, Prentice-Hall.

Robertson, I. and Downs, S. (1979) 'Learning and the prediction of performance: development of trainability testing in the United Kingdom', *Journal of Applied Psychology*, **64**(1), 42-50.

Robertson, I. and Kandola, R (1982) 'Work sample tests: Validity, adverse impat and applicant reactions', *Journal of Occupational Psychology*, **55**, 171-83.

Salvia, J. and Ysseldyke, J.E. (1998) *Assessment (7th Ed.)*. Boston, Houghton Mifflin Company.

Sambell, K., McDowell, E. and Brown, S. (1997) '"But is it fair?": An Eploratory Study of Student Perceptions of the Consequential Validity of Assessment', *Studies in Educational Evaluation*, **23**(4), 349-71.

Samsom, G.E., Grane, M.E., Weinstein, T. and Walberg, H.J. (1984) 'Academic and Occupational Performance: A Quantitative Synthesis', *American Educational Research Journal*, **21**(2), 311-21.

Schofield, J.W. (1993) 'Increasing the Generalizability of Qualitative Research' in Hammersley, M (ed) *Educational Research - current issues*. London. The Open University.

Shavelson, R., Baxter, G. and Pine, J. (1992) 'Performance Assessment: Political Rhetoric and Measurement Reality', *Educational Researcher*, **21**(4), 22-27.

Shepard, L. A. (1991) 'Psychometricians' beliefs about learning', *Educational Researcher*, **20**(7).

Shepard, L. A. (1993) 'Evaluating Test Validity', *Review of Research in Education*, **19**, 404-50.

Smith, M. L. and Shepard, L. (1988) Kindergarten readiness and retention. A qualitative study of teachers' beliefs and practices. *American Educational Research Journal*, **25**(3), 307-33.

Stake, R. E. (1978) 'The case-study method in social inquiry', *Educational Researcher*, **7**, 5-8.

TDT (1999) *Review of Royal Engineers' Clerks of Works Employment and Training*. Chatham, TDT.

Tittle, C. K. (1989) 'Validity: Whose Construction is it in the Teaching and Learning Context?', *Educational Measurement: Issues and Practice*, **8**(1), 6 - 15.

Tolley, G. (1989) *Learning and Assessment*, in Murphy, P. and Moon, R. (eds.) *Developments in Learning and Assessment*. London, Hodder & Stoughton.

Usher, R. and Bryant, I. (1987) 'Re-examining the theory-practice relationship in continuing professional education', *Studies in Higher Education*, **12**(2), 201-212.

Vincenti, W. G. (1993) *What Engineers Know and how They know it – Analytical Studies from Aeronautical History*. London, John Hopkins Press.

Williams, I.C. and Boreham, N.C. (1971) *The Predictive Value of CSE Grades in Further Education (Schools Council Examination Bulletin 24)*. London, Evans/Methuen Educational.

Wolf, A. (1995) *Competence-based Assessment*. Buckingham, OU Press.

Wolf, A. (1996) 'Vocational Assessment' in Goldstein, H. and Lewis, T. (eds) *Assessment: Problems, Developments and Statistical Issues - A volume of expert contributions*. Chichester, John Wiley & Sons.

Wolf, A., Silver, R., Collins, P. and Tambini, D. (1994) *Assessing Broad Skills. Final Report to the Employment Department*. London, Institute of Education.

Wood, R. (1991) *Assesment and Testing. A Survey of Research*. Cambridge, Cambridge University Press.

Wood, R., Johnson, C., Blinkthorn, S., Anderson, S. and Hall, J. (1989) *Boring, Blanching and Backtacking: Assessing Performance in the Work Place*. Sheffield, Training Agency.

Woods, P. (1996) 'Qualitative Research' in *Project Resources, Educational Research in Action, E835 Study Guide, Part 2.* London, Open University.

Wragg, E.C. (1994) *Conducting and Analysing Interviews* in Bennett, N., Glatter, R. and Levacic, R. (eds) *Improving Educational Management through research and consultancy.* London, Paul Chapman Publishing.

Yin, R.K. (1994) 'Designing Single- and Multiple-Case Studies' Bennett, N., Glatter, R. and Levacic, R. (eds) *Improving Educational Management through research and consultancy.* London, Paul Chapman Publishing.

Youngman, M.B., (1994) Designing and Using Questionnaires in Bennett, N., Glatter, R. and Levacic, R. (eds) *Improving Educational Management through research and consultancy.* London, Paul Chapman Publishing.

APPENDICES

Appendix A

Doctorate Research - Assessment in Engineering

Ethical Statement - Staff & Students

Introduction

Mr Brooks, the SLec(M), will be conducting research into the assessment of engineers over the next four years in order to obtain a Doctorate in Education from the Open University. The conclusion of the research will be a dissertation entitled '*Training the Military Engineer - Assessment and its Validity*' which will be completed in September 2000.

Research Ethics

The research will be conducted in accordance with the '*Ethical Guidelines for Educational Research*' adopted by the British Educational Research Association in 1992, a copy of which can be obtained from Mr Brooks.

The aim of the research is to examine the methods of assessment used throughout the Clerks of Works courses and how they are used to provide the quality of engineer required by the Royal Engineers. They will also be looked at with regard to how they contribute to the award of the BTEC HND.

All participants will be asked for their consent before any research is conducted and they have the right to withdraw from the study at any time. Throughout the research the nature of your responses will be known only to Mr Brooks and all responses will be treated in the strictest of confidence.

All participants will have the right to be informed of the likely publication of any findings that could have potential consequences for them and they will be given access to such documents. Should any participant withdraw from the study at any time, all reference to their individual responses will be removed from any publication.

Format of the Research

For staff, the research will take the form of interviews with individual staff and review of course documents, such as examinations, projects, marking schemes, course syllabi etc. For students, the research will take the form of student questionnaires and interviews with individual students. At each stage the exact nature of the research will be explained and participants will again be assured that they will not be identified with their responses.

DAJ BROOKS
SLec(M)
Ext 2398

Appendix B

Short Extract from Interview Transcript - L8

Interviewer: D A J Brooks (DB)

Interviewees: Military Lecturer (Mechanical) – (L8)

Date: 27 April 1999

Extract from Transcript:

DB: What do you feel the main aims of the Clerks of Works Course are?

L8: To produce somebody who is not going to be embarrassed when they get out there and has to do something. They are going to have all the skills to be able to carry out their job, and the Clerks of Works job out there is fast, so it is making them grow up into all-round individuals with an engineering bent.

DB: In producing this individual what knowledge do you think they need to have to be able to do their job?

L8: They need a fundamental background, a basic foundation to build upon. But engineering is so big they can't do everything so they have got to know how to build on that basic. The other thing is to know their limitations as well.

DB: When you say basic foundation, what sort of areas are you talking about?

L8: Generally, physics, how the world works, maths, and then a grounding of other subjects. How does water flow in pipes – the full in-depth of it is knowing there are different ways that it flows. Like a bigger pipe will flow differently to a smaller pipe. So a basis to stand it all on.

DB: Speaking of the water – what modules, what engineering knowledge are you responsible for?

L8: The hydraulics and water supply, including water treatment, pipes, pumps, hydrostatics, storage, air conditioning and psychometric processes through plant design and controls. I also do the small modules, climatology and medical gases.

DB: How is it defined what you are going to teach in the modules?

L8: Somewhere there is a set of training objectives that are probably very old by now, about 1989 I think. Basically, it is inherited from what was here before. Certain lesson plans set out things to be covered and I tried to tie those in with the BTEC Units. The course has been inherited and has now grown.

DB: In developing the course, what have you inputted and why?

L8: I looked at things that I didn't understand in the lesson, or seemed to have been sketched over when I did the course. So I learnt about them and tried to put them over in a different way. Also things that we have been doing wrong or we have not been doing, the fundamentals have been put in. Looking at the BTEC stuff and saying "we don't do that", finding out about it and putting it in. Saying that bit there needs adding to or bending to about 90 degrees and then it will fit it.

DB: Are you saying that you are developing the course from your own experience, that you are using your own judgement?

L8: Yes – especially on the water side, I did quite a bit of water when I first left the course and realized that on the course they taught a lot of things about pipe sizing. They taught us about a 100mm pipe, 90mm pipe, and 50mm pipe. But on the ground you have the 4", and that's all. So we do the pipe sizing still, because it is one of the fundamentals, the basics, but I ask 'what happens when you have only got 4" pipe?' How does this affect things? You've done this ideal thing, how do you work back? That's the sort of thing we put in.

DB: You spoke earlier about skills. As well as teaching the fundamental procedures, are you developing anything else in the students?

L8: Yes, an analysis of whether your resources are right or wrong. Yes, we have done this pipe-sizing exercise but is that practical, is it real, is it possible? You have to think about realism, that you have the basics and then you have to apply them to real life.

DB: Another thing that is linked to this perhaps is the military idea of working under pressure. It has been mentioned to me a number of times that they are learning to work under pressure. What is your view on that?

L8: I think that if you need to work under pressure you can. Whether you need to learn to work under pressure or not, I don't know. Whether it is "we worked every hour on the course, so we should be able to do it now". The people we turn off are well motivated and have got the personal skills that if they need to do it, they would do it, whether they have done it before or not. If the situation arose that they had to work under pressure they would be able to do it, I don't think it is something you can teach them, it is a personal thing. All that you can hope to get across to them is that if they need to do it, they will do it. It's part of growing up.

DB: And what about the confidence that you mentioned, that they start the course and they develop confidence. Is that part of the course itself?

L8: Yes, it is one of the things that I try and do. When we mark the course, we often find that they brush over some things. So we say, "it's down to you to do it, have the confidence to do it". But they don't really get the mega confidence boost until the final attachment when they go away. They come back from there 10ft tall, but before that it is difficult to give that confidence boost because they have got nothing to compare themselves against, they have only got the other people in the class.

...

DB: So if someone passes the two exams, they have been taught and they have been assessed. What does that pass mark tell you about the change in them? What is it telling you?

L8: So they have passed the test, the exams, they have done all the homeworks, got satisfactory marks on those, and the projects as well.

DB: Just deal with the exams.

L8: The exams show that they have understood what has been taught them, or what they should have learnt through directed reading, or whatever. And that they can relay that information back, I suppose, in the exams.

DB: And what about the project?

L8: The project is showing that they can apply that knowledge, they can apply that understanding. The exam is testing that they understand it and the project is showing that they can apply that understanding. I understand how a jet engine works, but I can't go and make one work.

DB: If we can go back to three things. You talked about training objectives, BTEC and you and your predecessors' views of the content of the modules. So when you say an understanding of an application of the knowledge, exactly which of those three areas of knowledge are you referring to?

L8: I think the BTEC points to an understanding. I think the army needs to apply that understanding.

DB: And where do your views, the views that you mentioned earlier on, fit in?

L8: I think that the major thing is not to go out there and be embarrassed or let anybody down. So I think so the understanding and applying the knowledge part is important. And that is more important if you are going to be the guy out there and they say, right, we need this water across that hill. Yes, you understand how to do flowing pipes, you understand all about Reynolds number and all the rest of it. But if you can't apply that and make the water go over the hill, you are

no use to anybody. BTEC probably look more at the other side, simply "does he understand what is going on?"

DB: And the application bit, in a class situation, you are teaching the general knowledge, how do you think they learn the application side of it?

L8: Hopefully all lectures put across the application as well to a certain extent. We do this "yes, we can size this pipe to 13.35 mm" but you don't get 13.35 mm pipe you get 15 mm pipe. Hopefully I apply my experiences of when I have done pipes and water and whatever. And the practicalities of that, you put in and derive what you are going to do. How do you apply this knowledge and hopefully it comes across a lot in the lectures. It is like a lot of things; it is down to general intelligence. You like to think that the people on the course have got the intelligence. That they know more than "I can do pipe sizing and if that is the pipe it comes out at then that is the pipe it comes out at." They should have more integrity than that and be able to put it in practically. Whether you can teach them that, I don't know. You can put ideas across but everybody has different views. Straight hydraulic gradients on a pipeline is an evenings thing and each student gives a different solution to it. Because they have each applied it differently. All you can do is to get them not to miss things. When you are applying this make sure you consider every possibility.

DB: Really then you are trying to teach them it and hoping that they are going to learn to develop their own style and the project is actually going to assess whether they have achieved that or not?

L8: Yes, to a certain extent. But it is something that they are not going to be able to do properly. They can do the project now, get a mark for it, come back in three years time when they have done it and do the project and they would do it a totally different way again. Because their knowledge is a thing that you can't give, ok two years is a long time, but you can't get them to go out and practice all the things.

...

DB: Throughout the course they have been trained to do a specific job as an engineer, clerks of works engineer. How well do you think the course as a whole, including the attachment, prepares them for the future job?

L8: I think it must do it very well, because it works. For individuals it would probably work to a different degree. Because we know that people aren't all the same. But I think it must do the job. It must work. I think we do prepare people well for what goes on out there.

DB: And do you think the way in which they are assessed means that you can say that "yes, they are going to be good engineers"?

L8: With the exams. Personally I don't mind exams, I can do exams, I don't get stressed out, I sleep the night before an exam. I know that some people get stressed out and don't give their best in an exam. So whether an exam is the best way of testing someone's knowledge, I'm not sure, because I have known people who have failed exams who would have made a good Clerks of Works but they have failed the course because of the exams. There was a lad on our course who couldn't do exams, and for nights before he wouldn't sleep. He was round at my house, we would do revision together and he wouldn't sleep. And when it came to the exams he would be physically drained before he started and he would get 10%. He just couldn't write anything. His practicality was absolutely brilliant. He could do it in the class. Practical-wise he had the experience of doing all sorts in the class. He was back-coursed and came top of the course behind us, because he did some exams again that he had done once, was taught again. He had more confidence and passed the exams and his confidence grew and grew and grew. He came top of the course behind us and he is now in Brunei. So I think it is the projects that really count.

DB: So these exams, did they come prior to projects so that he didn't have an opportunity to get to do any projects first time round?

L8: When we went through we didn't do mini-projects along the way. We did the major project at the end and he wouldn't have done the projects to have brought all that out. It would appear to me that the exams shouldn't be 90% and the others 10%. It should be more 50 – 50 or even 40 – 60 in favour of projects.

DB: Who is making that judgement?

L8: The thing has been inherited down the line. But each Senior Military Lecturer tweaks it themselves. I think that the ultimate responsibility is Pat's but he does discuss it with Tony and I.

DB: And that would be based on your feelings on how important exams are in relation to projects.

L8: Yes. We have brought exams down in the weightings and projects up.

DB: And is that acceptable to the Army and BTEC?

L8: Yes BTEC like us doing that.

DB: And what about the Army?

L8: As long as we are turning out people that can do the job, that's ok.

DB: And this is my final question for this interview, do you feel that you can actually say that the course produces a good product, a good engineer?

L8: Yes, I don't think that anybody gets through that shouldn't. I think that anybody that gets through is more than capable of doing the job the Army wants him to do. But the other way round is that maybe people who don't get through should do or could do. It has always worked in the past and maybe out of the hundreds that have passed, probably there are one or two that shouldn't for one reason or another. So I think yes, it works. What we are putting out is good but like all systems it is not infallible. You get people who fail but perhaps should pass, however I, personally, don't think anybody passes that should fail.

Appendix C

Doctorate in Education Research Study
Questionnaire 1
Clerk of Works Course Initial Questionnaire

Instructions

Please read each question carefully and answer by indicating a single score or by ticking the boxes you think reflect your views. In some cases you will probably need to tick more than one box. Please give open, honest replies as the information you supply will be used to carry out further research.

Questions

1. How appropriate do you think the following types of assessment are at assessing your ability to perform as engineers?

	Least Appropriate				Most Appropriate
	1	2	3	4	5
Homeworks **(HW)**	1	2	3	4	5
Lab Reports **(LR)**	1	2	3	4	5
Written Exams**(WE)**	1	2	3	4	5
Presentations **(P)**	1	2	3	4	5
Design Projects**(DP)**	1	2	3	4	5

2. What do you see as the main purpose for these assessments?

	HW	**LR**	**WE**	**P**	**DP**
To measure your understanding of the subject matter					
To predict how good you will perform as an engineer					
To test your memory skills					
To test your problem-solving skills					
To maintain the standards set by BTEC					

	HW	LR	WE	P	DP
To maintain the standards set by the Military Sponsor					
To assess your personal qualities: motivation;determination; working under pressure;etc.					
To give you and the lecturer feedback on your progress					
To give the lecturer feedback to develop module content					

3. Is it easy to understand what it is you are being asked to do in an assessment?

	Yes	No
Is it clear what is required in written lab reports?		
Are some exam questions presented in a confusing manner?		
When exam questions are related directly to engineering topics does this make them less confusing?		
Are the output requirements for projects made clear?		
Is it clear what criteria are being used to assess the projects?		
Is it clear what you need to do for a presentation?		
Is it clear what criteria are being used to assess a presentation?		

4. Do you think any of the following qualities are being assessed?

	Formally		Informally	
	Yes	No	Yes	No
Working under pressure				
Communication skills				
Motivation				
Determination				
Working as part of a team				
Creativity				
Initiative				

5. What do you think the course assessments are really testing or measuring?

	HW	LR	WE	P	DP
The assessment tests what you have been taught in class					
The assessment predicts you are very likely to pass the whole course					
The assessment predicts you are likely to be good in your future job					
The assessment measures your understanding in relation to the military training objectives					
The assessment measures your understanding in relation to the BTEC objectives					
The assessment is used by the lecturers to develop the subject matter					
The assessment is used by the lecturers to develop the teaching methods					
The assessment is used by the lecturers to improve future assessment methods					

6. At the end of the course do you think your potential to perform well as a Clerk of Works will have been fully and appropriately assessed?

Definitely Yes		Not Sure		Definitely No
1	2	3	4	5

7. On what basis do you make this judgement?

	YES	NO
Your knowledge of what a Clerk of Works does		
Discussion with the lecturers about what a Clerk of Works does		
Discussion with the lecturers about how you have performed on the course		
The nature of the course content		
The nature of the course assessments		
The award of the HND		
Other not listed above:		

Thank you for completing this questionnaire.

Please return the completed questionnaire to Mr D Brooks, SLec(M), by the close of play on ...

Appendix D

Doctorate in Education Research Study
Questionnaire 2
Clerk of Works End of Course Questionnaire

Instructions

Please read each question carefully and answer by indicating a single score or by ticking the boxes you think reflect your views. In some cases you will probably need to tick more than one box. Please give open, honest replies as the information you supply will be used to carry out further research.

Questions

1. How appropriate do you think the following types of assessment are at assessing your ability to perform as engineers?

	Least Appropriate				Most Appropriate
	1	2	3	4	5
Homeworks (**HW**)	1	2	3	4	5
Lab Reports (**LR**)	1	2	3	4	5
Written Exams(**WE**)	1	2	3	4	5
Presentations (**P**)	1	2	3	4	5
Design Projects(**DP**)	1	2	3	4	5

2. What do *you* see as the main purpose for these assessments?

	HW	**LR**	**WE**	**P**	**DP**
To measure your understanding of the subject matter					
To predict how good you will perform as an engineer					
To test your memory skills					
To test your problem-solving skills					
To maintain the standards set by BTEC					

	HW	LR	WE	P	DP
To maintain the standards set by the Military Sponsor					
To assess your personal qualities: motivation;determination; working under pressure;etc.					
To give you and the lecturer feedback on your progress					
To give the lecturer feedback to develop module content					

3. Is it easy to understand what it is you are being asked to do in an assessment?

	Yes	No
Is it clear what is required in written lab reports?		
Are some exam questions presented in a confusing manner?		
When exam questions are related directly to engineering topics does this make them less confusing?		
Are the output requirements for projects made clear?		
Is it clear what criteria are being used to assess the projects?		
Is it clear what you need to do for a presentation?		
Is it clear what criteria are being used to assess a presentation?		

4. Do you think any of the following qualities are being assessed?

	Formally		Informally	
	Yes	No	Yes	No
Working under pressure				
Communication skills				
Motivation				
Determination				
Working as part of a team				
Creativity				
Initiative				

5. What do you think the course assessments are really testing or measuring?

	HW	LR	WE	P	DP
The assessment tests what you have been taught in class					
The assessment predicts you are very likely to pass the whole course					
The assessment predicts you are likely to be good in your future job					
The assessment measures your understanding in relation to the military training objectives					
The assessment measures your understanding in relation to the BTEC objectives					
The assessment is used by the lecturers to develop the subject matter					
The assessment is used by the lecturers to develop the teaching methods					
The assessment is used by the lecturers to improve future assessment methods					

6. In what ways do you think the course has improved your ability to work as an engineer?

	Poor			Good	
Increased knowledge of foundation subjects	1	2	3	4	5
Increased knowledge of engineering subjects	1	2	3	4	5
Increased understanding of engineering concepts	1	2	3	4	5
Increased ability to solve problems	1	2	3	4	5
Increased ability to communicate ideas to others	1	2	3	4	5
Increased creative ability	1	2	3	4	5
Increased ability to work under pressure	1	2	3	4	5
Increased ability to work as part of a team	1	2	3	4	5

7. How do you think the above were assessed during the course?

	HW	LR	WE	P	DP
Increased knowledge of foundation subjects					
Increased knowledge of engineering subjects					
Increased understanding of engineering concepts					
Increased ability to solve problems					
Increased ability to communicate ideas to others					
Increased creative ability					
Increased ability to work under pressure					
Increased ability to work as part of a team					

8. What abilities and/or qualities do you think the Final Project assessed?

	Yes	No
Your knowledge of the foundation subjects		
Your knowledge of engineering subjects		
Your use of engineering procedures to solve problems		
Your use of engineering concepts to solve problems		
Your creative approach to problem solving		
Use of your initiative		
Your ability to work under pressure		
Your ability to work as part of a team		
Your ability to communicate your ideas in writing		
Your ability to communicate your ideas verbally		

9. Which of the following do you consider to be knowledge or abilites you need to have to be a competent engineer?

	Yes	No
1. Knowledge of the foundation subjects		
2. Knowledge of engineering subjects		
3. To be able to use engineering procedures to solve problems		
4. To be able to use engineering concepts to solve problems		

5. To have a creative approach to problem solving		
6. To be able to use your initiative		
7. To be able to work under pressure		
8. To be able to work as part of a team		
9. To be able to communicate your ideas in writing		
10. To be able to communicate your ideas verbally		

10. Which five of those listed above do you consider to be the most important? List the numbers in order of importance.

11. Considering questions 8, 9 & 10, do *you* think the Final Project gave you the opportunity to show how you will perform in your future job?

Definitely No		Not Sure		Definitely Yes
1	2	3	4	5

12. On what other factors do you make this judgement?

	YES	NO
Knowledge of the job		
Discussions with the lecturers		
Feedback from previous course assessments		
The content of the Final Project		
Feedback from the Final Project		
The grade you were given for the Final Project		
Your own perception of how you approached the task		
Your own perception of your solution to the task		

Others not listed above:

13. Do you think the course has successfully assessed your potential to perform well in your future job?

Definitely Yes		Not Sure		Definitely No
1	2	3	4	5

14. On what basis do you make this judgement?

	YES	NO
Knowledge of the job		
Discussion with the lecturers		
The nature of the course content		
The nature of the course assessments		
The award of the HND		

Others not listed above:

Thank you for completing this questionnaire.

Please return the completed questionnaire to Mr D Brooks, SLec(M), at the earliest opportunity and no later than

Appendix E

Student Interview Schedule

These are a selection of questions that guided my interviews with the students but I was prepared to follow any avenue that the students opened up.

What do you think is the purpose of the foundation phase of the course?
What do you think you learned from this phase?
How were you assessed during this phase?
What were you assessed on?
Do you think any of your skills, attitudes, personal qualities were being assessed?
What do you think was the purpose of the exams/labs/homeworks?
What does passing this phase say about you as a student?

What do you think you learned during the engineering phase?
How were you assessed? What methods were used?
What do you think was the most appropriate method of assessment?
Why do you think that?
What do you think was the least appropriate?
What do you think you were being tested on in exams/projects/presentations?
What was their purpose?
Who designs the exams and projects?
What skills were you being encouraged to develop?

Did you use anything you learned in the foundation phase later in the course?
What and how did you use it?

Does the course assess your future potential to do the job of a Clerk of Works?
How does it do this?
On what do you base your judgement?

What do you think you need to know to be a good Clerk of Works engineer?
What do you think you need to be able to do to be a good Clerk of Works engineer?
On what basis do you make this judgement?

What are the main things you have learned from doing the course?
In what ways are you different from when you started?

Is there anything else you would like to say about the course?

Thank you very much for your time.

Appendix F

Course Timetables

No 50 Clerks of Works (Mechanical) Course
7 January 1998 - 10 December 1999

Week no	Sponsor	Subject				
		Mon	Tue	Wed	Thu	Fri
1	E&M			Course muster & admin		RETD2
2	S&C	Basic computing, MSDOS, Maths revision, Engineering Science				
3	S&C	Computer applications, Engineering Science, Algebra				
4	S&C	Calorimetry & Heat Transfer, Algebra, Gas Laws				
5	S&C	Computer applications, Algebra, Graphs				
6	S&C	Computer programming, Geometry, Maths exam 1				
7	S&C	Velocity & Acceleration, Forces, Trigonometry, Computer applications				
8	S&C	Moments, Levers & Beams, Trigonometry, Frameworks				
9	S&C	Statistics, Algebra, Mechanics exam				
10	S&C	Maths exam 2, Calculus, Graphs, Angular motion, Friction, Projectiles				
11	S&C	Work, Power, Energy, Calculus, Simple Machines, Newtons Laws				
12	S&C	Simple machines, Algebra, Centrifugal Force			Properties of Materials	
13	S&C	Dynamics exam, Calculus			Calculus, Harmonics	
14	S&C	Mechanics practicals, Phase test			Easter leave	
15	S&C	Easter leave				
16	S&C	Thin cylinders, Trigonometry, Shear Force & Bending Moments				
17	S&C	Torsion Theory, Trigonometry				
18	S&C	Bank holiday	Metallurgy, Heat Treatment, Plastics			
19	S&C	Differential equations, Applications to Engineering				
20	S&C	Fourier Series, Laplace Transformation Applcations, Controls, Maths exam				
21	S&C	Bank holiday	Materials testing, Jointing methods			
22	Cmd	SNCOs Course				
23	Cmd					
24	Cmd					
25	S&C	Control Engineering				
26	CEW	Project Management				
27	E&M	Thermodynamics				
28	E&M	Thermodynamics				
29	E&M	E&M Drawing Office				
30	E&M	Compressed air		Energy Management		
31	E&M	Climatology				
32	E&M	Water Treatment				
33	Lve	Summer Block Leave				
34	Lve					
35	Lve					
36	E&M	Heating			Refrigeration	
37	E&M	Refrigeration		Heating		
38	E&M	Heating			Refrigeration	
39	E&M	Refrigeration		Heating		
40	E&M	Refrigeration				Exam
41	E&M	Heating			Exam	Exam
42	E&M	Boilerhouse technology		Air conditioning		

43	E&M	Air conditioning		Boilerhouse technology		
44	E&M	Boilerhouse technology		Air conditioning		
45	E&M	Air conditioning		Boilerhouse technology		
46	E&M	Boilerhouse technology		Air conditioning		
47	E&M	Air conditioning		Boilerhouse technology		
48	E&M	Boilerhouse technology			Exam	Exam
49	E&M	Air conditioning			Exam	Exam
50	E&M	CAE project				
51	Lve	Christmas Block Leave				
52	Lve					
53	E&M	Water supply			IC Engines	
54	E&M	IC Engines			Water supply	
55	CEW	Route recce			IC Engines Practical	
56	U&E	IC Engines Practical			Water supply	
57	E&M	Water supply				
58	E&M	IC Engines			Exam	Exam
59	U&E	POL				Water supply
60	E&M	Water supply				
61	E&M	AP PET				
62	E&M	Water supply			Exam	Exam
63	E&M	Water Supply Project				
64	E&M					
65	E&M	PPM			Easter Leave	
66	E&M	Easter Leave			Noise & Vibration	
67	E&M	LPG	Vertical Transportation		Medical gases	
68	E&M	Electrical Engineering				
69	E&M					
70	E&M					
71	E&M					
72	E&M					
73	E&M					
74	E&M	Bank Holiday		Spec writing		
75	S&C	Control Engineering 2				
76	CEW	Contract Procedures				
77	CEW	Contract Procedures				
78	CEW	Mech Trg Wg				
79	E&M	Ex Leading Edge				
80	E&M	Integrated Final Design Project				
81	E&M					
82	E&M					
83	E&M					
84	U&E	Machine Shop				
85	Lve	Summer Block Leave				
86	Lve					
87	Lve					
88	U&E	Welding				
89	CEW	RE SSS				
90	CEW	Building Technology				
91	CEW	Building Technology				
92	E&M	LPO Course			MWF	
93	U&E	Blacksmiths		P&D/Chippy	Power Hydraulics	

94	U&E	P&P		Fabrication		
95	Att	Civilian Industrial Attachment				
96	Att					
97	Att					
98	Att					
99	Att					
100	Attt					
101	E&M	Administration, End of Course Presentation				

No 44 Clerks of Works (Electrical) Course
18 May 1998 - 25 Feb 2000

Week no	Sponsor	Subject	Week no	Sponsor	Subject
1	E&M	Course Intro, Maths 1	48	E&M	Illumination
2	S&C	Maths 1	49	E&M	Illumination
3	S&C	Maths 1	50	E&M	Electrical Installation
4	S&C	Maths 1	51	E&M	Electrical Installation
5	S&C	Maths 1	52	E&M	Electrical Installation
6	S&C	Maths 1	53	E&M	Electrical Installation
7	S&C	Engineering Science	54	E&M	Electrical Installation
8	S&C	Engineering Science	55	CEW	Surveying
9	S&C	Engineering Science	56	E&M	AC Machines
10	S&C	Engineering Science	57	E&M	AC Machines
11	S&C	Maths 2/Computing	58	E&M	AC Machines
12	S&C	Maths 2/Computing	59	E&M	Ex Leading Edge
13	S&C	Maths 2/Computing	60	E&M	AC Machines
14	Lve	Summer Block Leave	61	E&M	AC Machines
15	Lve		62	E&M	Transmission & Distribution
16	Lve		63	E&M	
17	E&M	DC Technology	64	E&M	
18	E&M	DC Technology	65	Lve	Summer Block Leave
19	E&M	DC Technology	66	Lve	
20	E&M	AC Fundamentals	67	Lve	
21	E&M	AC Fundamentals	68	E&M	Trans & Distribution
22	E&M	AC Fundamentals	69	E&M	LPO/Engr Res
23	CEW	SNCOs Course	70	CEW	RE SSS
24	CEW		71	E&M	Switchgear & Protection
25	CEW		72	E&M	
26	S&C	Electronics/Controls	73	E&M	
27	S&C	Electronics/Controls	74	E&M	AP Elect Course
28	S&C	Electronics/Controls	75	E&M	AP Elect Course
29	S&C	Electronics/Controls	76	E&M	Switchgear & Protection
30	S&C	Electronics/Controls	77	E&M	
31	E&M	Engineering Drawing	78	E&M	
32	Lve	Christmas Block	79	E&M	Integrated Final Design Project
33	Lve	Leave	80	E&M	
34	E&M	AC Polyphase	81	E&M	
35	E&M	AC Polyphase	82	E&M	
36	E&M	AC Polyphase	83	E&M	
37	E&M	Mech Engineering	84	E&M	
38	E&M	Mech Engineering	85	Lve	Christmas Block Leave
39	E&M	Mech Engineering	86	Lve	
40	E&M	Mech Engineering	87	CEW	Cascade Planning
41	E&M	Mech Engineering	88	E&M	Civilian Industrial Attachment
42	CEW	Building Technology	89	E&M	
43	CEW	Building Technology	90	E&M	
44	CEW	Quantity Surveying	91	E&M	
45	CEW	Quantity Surveying	92	E&M	
46	E&M	Illumination	93	E&M	Administration, End of Course Presentation
47	Lve	Easter Leave			

Appendix G

Water Supply Module - Main Project

HIRTA GARRISON

MECHANICAL & ELECTRICAL SERVICES

PROJECT DIRECTIVE

References:

A. Project Brief.

B. Drawing T/CWM/1249 - Hirta Hospital Site Plan.

C. Location Map of Sub-stations and Water Source.

GENERAL

1. The Hirta Garrison design is now nearing completion. However, due to shortfalls in manpower, Military Works Force (MWF) have not been able to consider the water supply or sewage elements of the project.

2. The existing potable water supply for the garrison is stored in 3 x 160 m glass reinforced plastic (GRP) sectional tanks, each approximately 15 years old. They are located adjacent to the 174 ft trig point at the end of the Dry Burn. It is considered that the tanks have exceeded their economic life span and will be replaced during this phase of the project.

3. It is estimated that the boilerhouse will require 3.5 m^3 of potable water per day. Table 1, below, shows the analysis results for raw water samples taken at Gleann Mor.

Serial	Item	28 Jul 98	20 Feb 99
(a)	(b)	(c)	(d)
1	Colour	Clear	Brown
2	Odour	Slight	Nil
3	Turbidity	Slight	Turbid
4	'E' Coli per mil	956	243
5	Total Alkalinity	168	180

6	TDS	540	480
7	Hardness Ca	80	46
8	Hardness Mg	5	3
9	pH	5.4	6.4

Table 1. Physical ans chemical analysis of water at Gleann Mor.

REQUIREMENTS

4. You are to prepare a Design Report in three volumes, iaw RETD2, as follows:

a. Volume 1. Volume 1 will include the usual introduction format to the Design Report. The design report will include sufficient Appendices to cover the project directive/brief and location maps. The design report will be a stand alone methodological approach on how you would install the plant and equipment selected indicating, where necessary, any problems e.g. factors affecting the planned task. You should not include any appreciation in this volume unless it is pertinent in qualifying a particular statement. Volume 1 must be able to be read in conjunction with your working drawings.

b. Volume 2. Volume 2 will provide an appreciation of all your design work, the advantages and disadvantages of one system over another, any problems or limitations identified during your design assessment and the reasons why you are selecting one particular item of plant or equipment over another. Volume 2 will also include calculations where necessary to support an appreciation or to qualify a design element in Volume 1. Volume 2 will commence with a brief introduction and contents page followed by your appreciations and calculations in Appendices. Volume 2 must be able to be read in conjunction with your working drawings.

c. Volume 3. Volume 3 will contain full SKETCHES to support the design report and the appreciation.

5. The following points are to be considered in your report:

a. Water Supply. A detailed plan for the water supply which is to include:

(1) Development of the source.

(2) Selection and operation of the primary (main) pump sets. A single track metalled road will be designed by a CW© and laid to Gleann mor.

(3) A pipeline from the source to a storage and treatment site suitable for supplying both the hospital and the garrison.

(4) A water treatment plant.

(5) A pipeline from the storage to the hospital distribution system. The distribution system will terminate at all incoming stop cocks.

(6) A water supplied fire fighting installation to cover all buildings within the hospital grounds.

(7) A storage system to replace the 3 GRP garrison storage tanks.

b. Power. Design of the electrical power requirements for the water treatment plant and pumping statin, including the production of schematic diagrams showing the distribution board(s) containing cable sizes and fuse ratings or equivalent.

c. Sewage. Design of a suitable sewage plant to serve both the garrison and hospital population based on the numbers given in the project brief.

d. Stores Requirement. A costed stores list for the mechanical and electrical equipment installed in the water supply pipeline system as delivered to Benbecula. The sewage design is excluded from this requirement.

e. Costs. Provide a detailed estimate of the total civilian value of the works, using DOE/PSA schedules of rates, or similar, for similar work in Benbecula.

f. Inspection and Testing. An inspection, testing and commissioning procedure for the potable water system.

g. Works Programme. A works programme for the installation of the water supply equipment from the raw water source to the boundary fence of the hospital complex based on the strength of one fully established field troop, stating any specialised artisan tradesmen required.

h. Specification. The specification of all mechanical plant, services and works. You are to use National Building and Engineering Specifications as appropriate.

LIMITATIONS

6. All works, materials and specifications in your report are to comply with relevant statutes, British Standards and accepted good practice.

7. Your completed report is to be handed toWO2 Mechanical by ...

WO2 Mechanical
for Chief Lecturer

Hirta Garrison Project Marking Schedule

The total marks allocated for the project are 2272. These are for individual items in the report and design. The marks are broken down into areas, with a percentage of the total marks available as shown. This should give an indication of the time and effort required for each area.

Master Schedule

	% marks	Notes
Presentation and format	3	Including layout, grammar etc.
Summary	4	
Main body of text	15	
Raw water supply	14	
Source of treatment	12	
Water treatment plant	16	
Treatment to demand points	5	
Garrison storage	2	
Sewage disposal	2	
Fire	5	
Incidentals	8	Stores list, costs, testing, commissioning , Cascade
Sketches	14	
Total	100	

Itemised Schedule

Presentation and Format

Heading	Item	Mark
	Title	1
	Security Classification	1
	Date	1
	File Number	1
	Production location	1

	Unit	1
	Names & page numbers	8
Title sheet as above plus:	Distribution page	5
	Contents page	5
	Orders	1
	Originator	1
	Recipient	1
	Date issued	1
	Place issued	1
	Copy at Appendix A	1
	Where required	1
	Submitted to	1
	Submission date	1
	Reasons	1
	Limitations	1
References/Bibliography		10
Neatness and legibility		10
Spelling		10
Grammar and punctuation		10
Style	Continuity	10
	Sequence	10
	Sub total:	94

Summary

Heading	Item	Mark
Originality		25
Background	Where	5
	Why	5
Reference to orders	When	1
	Where	1
	Who	1
Requirement/Sequence		10
Special considerations		5

Design features		5
Stores		5
Manpower		5
Costs		5
Timings		5
Limitations (problems)		5
Conclusions		5
Recommendations		5
	Sub Total:	93

The Main Report

Heading	Item	Mark
Situation	History	10
	Current	10
	Foreseeable future	10
	Orders	5
Introduction		10
Aim		5
Requirement		5
Location		5
Scope		5
Total requirement		5
Fire fighting requirement		5
Reference to calculations		5
Treatment analysis		5
Chemicals for treatment		5
Storage/Reservoir of raw water		5
Fittings for storage		5
Storage at WTP		5
Fittings at WTP		5
Storage of potable water		5
Fittings for potable water		5
Pumps raw		5
Pumps WTP		5
Pumps distribution		5

Pumps fire fighting		5
Fresh water controls		5
Fire water controls		5
Electrical supply		10
Pipe material		5
Pipe jointing methods		5
Pipe anchorages		5
Pipe insulation/frost		5
Pipe marking		5
Break pressure problems		5
Syphonic problems		5
Instruments/meters/gauges		5
Civil work		5
Hydraulic gradient		10
Variations in pressure		5
Venting/drainage/scour		5
Selection of route to WTP		5
Velocity flow		5
Pumping timetable		10
Standby arrangements		5
Safety		5
Distribution in grounds		10
Distribution to sewage		10
Distribution to camp		10
Fire fighting standby facilities		5
Operating pressures		5
Reference to	Specification	5
	Stores list	5
	Installation costs	5
	Running costs	5
	Inspection & testing	5
	Commissioning	5
	Drawings	5
Additional comments		5
Conclusions		5

Recommendations		5
	Sub total:	350

Raw Water Supply

Heading	Item	Mark
Appreciation of requirement	General	20
	Manpower	5
	Vehicles	5
	Sewage	5
	System losses	5
	Fire fighting	10
Calculation of above		30
Water analysis appreciation		20
Initial treatment/chemicals		5
Methods of storage of untreated water		15
Level control		10
Overflow and drainage		10
Civil works		10
Pump selection (to WTP)		20
Pump installation		20
Power supply		10
Standby power supply		10
Pipework selection, material, method of jointing, configuration		20
Pipework calculations		10
Civil works, anchors etc.		20
Method of operation of pumps		20
Pressure gauges/meters		10
Installation/frost protection		5
	Sub Total:	295

Source to Water Treatment Plant (WTP)

Heading	Item	Mark
Pipeline material jointing		10
Layout of route appreciation		20
Frost protection/insulation		20
Additional pumping stations		20
Hydraulic gradients		20
Peak arrangements		20
Associated pressure applications		20
Air venting		10
Break pressure appreciation	Down hill	20
No flow hydraulic gradient	Down hill	20
Route of pipeline	Down hill	20
Velocity of flow	Down hill	5
Anchors and thrust blocks	Down hill	5
Valve pits, scour valves	Down hill	20
Control of flow	Down hill	20
Pipework calculations	Down hill	20
	Sub Total:	270

Water Treatment Plant

Heading	Item	Mark
Storage complete system		20
Plant selection		20
Plant control		20
Plant installation		20
Operating procedures		20
Flow diagrams (sketches)		10
Power supply		10
Standby power supply		10
Pipework design		20
Sterilization design		20
Storage of chemicals		15

Balance tank		5
Selection of pumps		20
Civil works		20
Layout of plant room		20
Size of plant room		20
Heating requirements		10
Meters and gauges		10
Drainage		20
Pipe jointing configuration		15
Pipe calculations		20
Compressor selection		10
	Sub Total:	355

WTP to Demand Points

Heading	Item	Mark
Distribution to hospital route		20
Distribution to garrison route		20
Termination in hospital buildings		10
Access into boiler house		5
Air Vent		5
Valve pits		10
Distribution to sewage equipment		10
Pipe calculation		20
Pipe material, jointing		10
Frost protection		10
	Sub Total:	120

Garrison Storage

Heading	Item	Mark
Complete design	Control	10
	Tank	10

	Calculations	10
	Civil works	10
	Materials	5
	Frost protection	5
	Sub Total:	50

Sewage

Heading	Item	Mark
Complete appreciation	Numbers	10
	Methods	10
	Equipment	10
	Routes	10
	Materials	10
	Control	10
	Disposal (final)	10
	Sub Total:	70

Fire Fighting

Heading	Item	Mark
Appreciation of requirements		20
Pump selection		10
Pump installation		10
Hydrant details		10
Pump operation		10
Pump control		10
Power supply		10
Standby power supply		10
Standby pumps		10
Meters and gauges		10
	Sub Total:	110

Incidentals

Heading	Item	Mark
Electrical	Power WTP	20
	Lighting WTP	10
Specifications of main plant items	Performance specifications	20
Stores list	Pipework	5
	Gaskets	5
	Flanges	5
	Insulation	5
	Valves	5
	Gauges	5
	Meters	5
	Tanks	5
	Tank fittings	5
	Pumps	5
	Controls	5
	Filters	5
	Marking tape	5
	Air Vents	5
	Bedding materials	5
Installation costs	Up to date	10
Annual running costs	Consumables	5
	Electricity	5
Inspection, testing, commissioning	Regulations	10
	Legislation	10
	Codes of practice	10
	Safety	10
	Is it reasonable	20
	Sub Total:	205

Drawings and Sketches

Headings	Item	Mark
Sewage treatment route		20

Fire fighting route		10
Anchors and supports	Guides, expansion devices	20
Raw water reservoir		20
WTP		20
WTP storage details		10
Hospital storage details		20
Pipe route	Up hill	20
	Down hill	20
	Hospital grounds	20
Air vent, scour valves, valve pits		20
Hydraulic gradients		20
WTP power supply		20
Supply system to garrison		20
	Sub Total:	260